UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays in Energy and Environmental Economics

Permalink
https://escholarship.org/uc/item/34p4q4sg

Author
Woerman, Matthew A

Publication Date
2019

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Essays in Energy and Environmental Economics

by

Matthew A. Woerman

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Meredith Fowlie, Chair
Professor Maximilian Auffhammer
Professor Lucas Davis

Spring 2019

Essays in Energy and Environmental Economics

Copyright 2019
by
Matthew A. Woerman

Abstract

Essays in Energy and Environmental Economics

by

Matthew A. Woerman

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Associate Professor Meredith Fowlie, Chair

Electricity is an essential input in the modern economy and a necessity of modern life in the developed world. However, the production of electricity and many of its end uses can be subject to market failures. In this dissertation I study two such market failures related to electricity: (i) power plants exercising market power in response to congestion of the electricity transmission grid; and (ii) electricity consumption for groundwater extraction, which imposes a spatial externality on neighbors. I finally develop new methods for designing experiments and performing power calculations for use with panel data, which is particularly relevant in electricity research that uses high-frequency data on electricity production or consumption.

In the first chapter, I examine how congestion of the electricity transmission grid affects the market in which a power plant competes and the market power exercised by that power plant. Economic theory tells us that market structure is the primary determinant of a firm's ability to exercise market power. However, it is challenging to empirically estimate the causal effect of market structure on market power because a firm rarely experiences exogenous variation in its market's structure. In this paper, I exploit a novel source of exogenous variation in market size within the Texas electricity market — congestion of electricity transmission lines due to ambient temperature shocks — to estimate the causal effect of market size on the exercise of market power. When transmission lines congest, this statewide market splits into smaller localized markets. I find that a 10% reduction in market size causes firms to more than double markups. The direction of this effect is consistent with a model of oligopoly competition in which firms set markups in response to residual demand, which is less elastic in a smaller market. My results imply that the markups induced by transmission congestion at high temperatures generate \$7.1–21.5 million of deadweight loss annually. These markups also create large transfers — \$2.1 billion per year — from consumers to producers, which raise important equity concerns.

In the second chapter, coauthored with Fiona Burlig and Louis Preonas, we examine an economically important electricity end use: groundwater extraction in California agriculture. Agricultural and water pumping demand for electricity comprises about 8 percent of California's total electricity usage. Groundwater pumping, a significant share of this usage, creates even more external costs than typical electricity consumption. When one farm pumps

groundwater, it lowers the level of the aquifer and increases pumping costs for neighboring farms—who now must pull groundwater a greater vertical distance. We study farmers' electricity usage for groundwater pumping and estimate the extent to which one farm's pumping behavior hurts neighboring farms. We assemble a novel dataset that combines (i) detailed data on farmers' electricity consumption, (ii) rich data from technical audits of these farmers' groundwater pumps, and (iii) publicly available measurements of groundwater depths in California aquifers. Using changes in agricultural electricity prices, we estimate farmers' pumping behavior to be far more price responsive in the medium-term than previously thought. However, we find no evidence that farmers respond to within-day price changes, suggesting limited opportunities for demand response from the agricultural sector. We then calculate the extent to which one farm's water pumping from a shared aquifer raises costs for neighboring farms. Our results suggest that the magnitude of the "pumping cost" spillover effect is fairly small compared to each farm's private cost of pumping groundwater.

In the third chapter, coauthored with Fiona Burlig and Louis Preonas, we answer the question: how should researchers design panel data experiments, particularly when using high-frequency data common in electricity research? We analytically derive the variance of panel estimators, informing power calculations in panel data settings. We generalize Frison and Pocock (1992) to fully arbitrary error structures, thereby extending McKenzie (2012) to allow for non-constant serial correlation. Using Monte Carlo simulations and real-world panel data, we demonstrate that failing to account for arbitrary serial correlation *ex ante* yields experiments that are incorrectly powered under proper inference. By contrast, our "serial-correlation-robust" power calculations achieve correctly powered experiments in both simulated and real data. We discuss the implications of these results, and introduce a new software package to facilitate proper power calculations in practice.

Contents

Contents **i**

List of Figures **v**

List of Tables **vii**

Acknowledgements **ix**

1 Market Size and Market Power: Evidence from the Texas Electricity Market **1**

1.1 Introduction . . . 1
1.2 Texas electricity market . . . 5
1.2.1 Real-time market . . . 6
1.2.2 Transmission congestion and local markets . . . 7
1.2.3 Power plants and firms . . . 9
1.3 Theoretical framework . . . 10
1.3.1 Optimal markup rule . . . 10
1.3.2 Comparative statics . . . 11
1.4 Data . . . 13
1.4.1 Offer curves . . . 13
1.4.2 Marginal cost . . . 14
1.4.3 Local market size . . . 15
1.4.4 Temperature . . . 16
1.4.5 Additional data . . . 16
1.5 Empirical strategy . . . 17
1.5.1 Prediction model . . . 18
1.5.2 Empirical specification . . . 21
1.6 Results . . . 23
1.6.1 First-stage results . . . 23
1.6.2 Reduced-form results . . . 24
1.6.3 Results for local market capacity . . . 25
1.6.4 Results for number of units or firms in local market . . . 26
1.6.5 Decomposition of effects . . . 27
1.6.6 Robustness checks for local controls and temperature matching . . . 28

1.6.7 Matched temperature or local temperature . . . 28
1.7 Welfare effects . . . 29
1.7.1 Consumer surplus and transfers to producers . . . 29
1.7.2 Climate change effects . . . 31
1.8 Conclusion . . . 32

2 Spatial Externalities in Groundwater Extraction: Evidence from California Agriculture 55
2.1 Introduction . . . 55
2.2 Background . . . 57
2.2.1 Groundwater as an open-access resource . . . 57
2.2.2 Agriculture in California . . . 58
2.2.3 Electricity for Pumping . . . 59
2.3 Data . . . 59
2.3.1 Electricity Data . . . 59
2.3.2 Pump Data . . . 60
2.3.3 California Water Data . . . 61
2.3.4 Groundwater Prices and Quantities . . . 62
2.4 Empirical Strategy . . . 63
2.4.1 Electricity Demand . . . 63
2.4.2 Groundwater Demand . . . 64
2.5 Results . . . 66
2.5.1 Electricity Demand . . . 66
2.5.2 Groundwater Demand . . . 67
2.6 Open-Access Externality . . . 69
2.7 Conclusion . . . 72

3 Panel Data and Experimental Design 85
3.1 Introduction . . . 85
3.2 Power calculations for panel data . . . 87
3.2.1 Serial-correlation-robust power calculations . . . 88
3.2.1.1 Accounting for serial correlation . . . 89
3.2.1.2 Monte Carlo simulations . . . 92
3.2.2 Sensitivities . . . 94
3.2.2.1 Short panels . . . 94
3.2.2.2 Alternative estimators . . . 94
3.2.2.3 Alternative data generating processes . . . 97
3.3 Applications to real-world data . . . 97
3.3.1 Bloom et al. (2015) data . . . 97
3.3.2 Pecan Street data . . . 99
3.3.3 Power calculations in real data . . . 99
3.4 Power calculations in practice . . . 100
3.4.1 Trading off units and time periods . . . 100
3.4.2 Power calculations in Stata . . . 102

3.4.2.1 Analytical DD power calculations . . . 102
3.4.2.2 Power calculations by simulation . . . 103
3.5 Conclusion . . . 104

References **115**

A Appendix: Market Size and Market Power: Evidence from the Texas Electricity Market **124**
A.1 Texas electricity market . . . 124
A.1.1 Bilateral contracts and day-ahead market . . . 124
A.1.2 Offer price mitigation . . . 125
A.2 Theoretical framework . . . 125
A.2.1 Optimal markup rule . . . 125
A.2.2 Comparative statics . . . 128
A.3 Data . . . 129
A.3.1 Marginal cost . . . 129
A.3.2 Temperature . . . 131
A.4 Results . . . 131
A.4.1 Decomposition of effects . . . 131
A.4.2 Robustness checks for local controls and temperature matching . . . 141
A.4.3 Matched temperature or local temperature . . . 157
A.4.4 Offer prices . . . 163
A.4.5 Heat rate . . . 169

B Appendix: Panel Data and Experimental Design **171**
B.1 Derivations and proofs . . . 171
B.1.1 Cross sectional experiments . . . 172
B.1.2 Panel experiments . . . 174
B.1.2.1 Independent error structure . . . 174
B.1.2.2 Serially correlated error structure . . . 177
B.1.2.3 Serially and cross-sectionally correlated error structure . . . 180
B.1.2.4 Collapsed dataset . . . 183
B.1.2.5 Analysis of covariance (ANCOVA) . . . 186
B.1.3 Arbitrary cross-sectional correlations . . . 193
B.1.4 Equivalence of alternative difference-in-difference estimators . . . 199
B.2 Figures in main text . . . 201
B.2.1 Simulated AR(1) data . . . 201
B.2.2 Bloom et al. (2015) data . . . 205
B.2.3 Pecan Street data . . . 208
B.2.4 Analytic power calculations . . . 209
B.3 Additional results . . . 210
B.3.1 Sensitivities with real data . . . 210
B.3.2 Tradeoffs between MDE and time periods: ANCOVA . . . 212
B.3.3 Cluster randomization in panel RCTs . . . 213

B.4 A practical guide to power calculations 216
B.4.1 Analytical power calculations 217
B.4.2 Simulation-based power calculations 225
B.4.3 Lack of (representative) pre-existing data 226
B.4.4 STATA package `pcpanel` 228
B.5 Estimation-related proofs 230
B.5.1 Recovering estimated parameters 230
B.5.2 Estimating MDE from residual-based parameters 232
B.5.3 Estimation with ANCOVA 242

List of Figures

1.1 Offer curves . . . 34
1.2 Local markets and locational marginal prices . . . 35
1.3 Units by fuel type . . . 36
1.4 Example supply offer curves . . . 37
1.5 Average supply offer curves and marginal costs . . . 38
1.6 Number of local markets by temperature . . . 39
1.7 Distribution of prediction errors . . . 40
1.8 Distance from power plants to weather stations . . . 41
1.9 Distribution of matched temperatures . . . 42
1.10 Temperature-response functions for market size measures . . . 43
1.11 Temperature-response functions for markups . . . 44
1.12 Number of local markets under climate change . . . 45

2.1 Average Marginal Electricity Prices . . . 74
2.2 PGE Agricultural Customers . . . 75
2.3 Framework for Calculating the Pumping Cost Externality . . . 76
2.4 Pumping Cost Externality: Densities for San Joaquín Valley . . . 77

3.1 Standard methods result in improperly powered experiments in AR(1) data . . . 105
3.2 Power in short panels – AR(1) data . . . 106
3.3 SCR outperforms traditional methods with alternative estimators . . . 107
3.4 Standard ANCOVA methods fail under serial correlation . . . 108
3.5 SCR outperforms traditional methods with alternative DGPs . . . 109
3.6 Power simulations for Bloom et al. (2015) data . . . 110
3.7 ANCOVA power calculations in real data . . . 111
3.8 Power simulations for Pecan Street data . . . 112
3.9 Analytical power calculations – daily Pecan Street dataset . . . 113
3.10 Analytical power calculations with increasing panel length . . . 114

A.1 Temperature-response functions for local market capacity by owner or unit type 138
A.2 Temperature-response functions for units in local market by owner or unit type 139
A.3 Temperature-response functions for firms in local market by unit type . . . 140
A.4 Robustness checks — Effects on markups using temperature bins or alternate quantities . . . 144

A.5 Temperature-response functions for local market capacity with other local controls . . . 145
A.6 Temperature-response functions for local market capacity with minimum distance . . . 146
A.7 Temperature-response functions for units in local market with other local controls . . . 147
A.8 Temperature-response functions for units in local market with minimum distance 148
A.9 Temperature-response functions for firms in local market with other local controls . . . 149
A.10 Temperature-response functions for firms in local market with minimum distance 150
A.11 Temperature-response functions for markup at 85% capacity with other local controls . . . 151
A.12 Temperature-response functions for markup at 85% capacity with minimum distance . . . 152
A.13 Temperature-response functions for markup at 75% capacity with other local controls . . . 153
A.14 Temperature-response functions for markup at 75% capacity with minimum distance . . . 154
A.15 Temperature-response functions for markup at 65% capacity with other local controls . . . 155
A.16 Temperature-response functions for markup at 65% capacity with minimum distance . . . 156
A.17 Temperature-response functions for market size measures using local temperature . . . 161
A.18 Temperature-response functions for markups using local temperature . . . 162
A.19 Temperature-response functions for offer prices . . . 168
A.20 Temperature-response function for heat rate . . . 170

B.1 Pecan Street data – Varying levels of aggregation . . . 209
B.2 Power in short panels – Real data . . . 211
B.3 Sensitivities to estimating equation – Real data . . . 212
B.4 Analytical power calculations with increasing panel length – ANCOVA . . . 213
B.5 Power calculations by simulation – Cluster randomization . . . 216
B.6 Actual vs. estimated parameters – AR(1) data . . . 221
B.7 Estimated parameters – Bloom et al. (2015) data . . . 223
B.8 Estimated parameters – Pecan Street data . . . 224

List of Tables

1.1 Summary statistics by unit type and firm 46
1.2 Summary of prediction models 47
1.3 Local market capacity — Effect on markups 48
1.4 Number of units in local market — Effect on markups 49
1.5 Number of firms in local market — Effect on markups 50
1.6 Local market capacity by owner and unit type — Effect on markups 51
1.7 Robustness to local controls — Effect of local market capacity on markups . . 52
1.8 Robustness to matching distance — Effect of local market capacity on markups 53
1.9 Temperature selection — Effects on markups 54

2.1 PGE Agricultural Tariffs 78
2.2 Summary Statistics – Electricity Data 79
2.3 Summary Statistics – Pump Tests and Groundwater Consumption 80
2.4 Estimated Demand Elasticities – Electricity 81
2.5 Estimated Demand Elasticities – Groundwater 82
2.6 Estimated Demand Elasticities – Groundwater 83
2.7 Deadweight Loss Calculations for June 2016 84

A.1 Units in local market by owner and type — Effect on markups 133
A.2 Firms in local market by unit type — Effect on markups 134
A.3 Local market capacity by owner and unit type — Effect on markups using temperature bins or alternate quantities 135
A.4 Units in local market by owner and type — Effect on markups using temperature bins or alternate quantities 136
A.5 Firms in local market by unit type — Effect on markups using temperature bins or alternate quantities 137
A.6 Robustness checks — Effect of units in local market on markups 142
A.7 Robustness checks — Effect of firms in local market on markups 143
A.8 Temperature selection — Effects on markups using temperature bins 158
A.9 Temperature selection — Effects on markups at 75% capacity 159
A.10 Temperature selection — Effects on markups at 65% capacity 160
A.11 Robustness to outcome measure — Effects on offer prices at 85% capacity . . 164
A.12 Robustness to outcome measure — Effects on offer prices using temperature bins 165

A.13 Robustness to outcome measure — Effects on offer prices at 75% capacity . . 166
A.14 Robustness to outcome measure — Effects on offer prices at 65% capacity . . 167

B.1 Summary statistics – Bloom et al. (2015) 206
B.2 Summary statistics – Pecan Street 208
B.3 Pecan Street Simulation Parameters 210

Acknowledgements

I am deeply grateful to my advisor, Meredith Fowlie. She has been a role model and has provided me with invaluable guidance since the start of my graduate studies at Berkeley. She has given me the freedom to pursue my own ideas and research, allowing me to learn from my mistakes and grow from these experiences, while still providing the support and direction to ensure that I succeed. She has taught me not only how to ask important research questions, but also how to answer those questions in a rigorous and convincing way.

I am also thankful for the support and guidance of Maximilian Auffhammer and Lucas Davis. Max taught me the importance of exogenous variation, and Lucas showed me the value of keeping research simple. I am further thankful for having been given the opportunity to be part of the Energy Institute at Haas. The EI faculty, staff, and fellow graduate students have all played an important role in my graduate education. I have learned more from sharing an office space and a lunch table than I ever thought possible.

Fiona Burlig and Louis Preonas taught me that, whether crowded around a whiteboard or flung around the world, research is more fun when you do it with your friends. Erin Kelley, Greg Lane, and Ed Rubin provided unending support and friendship that helped me through these six years. Patrick Baylis, Susanna Berkouwer, Josh Blonz, Judd Boomhower, Karl Dunkle Werner, Jimmy Gillan, Stephen Jarvis, John Loeser, Eva Lyubich, Dave McLaughlin, Andrew Stephens, Becca Taylor, and Derek Wolfson each contributed more to my success than they know. I cannot imagine a better group of classmates, coauthors, colleagues, and friends.

Most importantly, to my family, thank you for your endless love and support. You helped me through the difficult times and celebrated my victories. I will be forever grateful to each of you for all you have done. Amanda, we have been on an incredible journey together, and I am so excited to experience all that is still to come.

Chapter 1

Market Size and Market Power: Evidence from the Texas Electricity Market[1]

1.1 Introduction

Economists have long been concerned by a firm's ability to exercise market power because of the inefficiencies it imposes on that market and the economy more broadly. The primary consequence of market power is deadweight loss due to inefficient prices and levels of production (Harberger 1954), but additional consequences also exist. For example, firms that face less competition are slower to innovate (Aghion et al. 2005) and may provide lower quality (Gaynor and Town 2011). In fact, many recent macroeconomic trends, from a decrease in labor force participation to a slowdown in aggregate productivity, are consistent with increases in market power that have been observed globally in recent decades (De Loecker and Eeckhout 2017, 2018).

With such wide-reaching consequences, it is important to understand the factors that influence noncompetitive firm behavior. Economic theory tells us the primary factor is market structure (Cournot 1838), but it is challenging to empirically estimate the effect of market structure on market power. Both are equilibrium outcomes, making the causal nature of their relationship difficult to establish. In the absence of exogenous variation in

[1]I thank Meredith Fowlie, Lucas Davis, and Maximilian Auffhammer for their invaluable guidance and advice. I also thank Peter Berck, Susanna Berkouwer, Joshua Blonz, Severin Borenstein, Fiona Burlig, James Gillan, Nick Hagerty, Solomon Hsiang, Larry Karp, Erin Kelley, Gregory Lane, John Loeser, Jeffrey Perloff, Louis Preonas, Gordon Rausser, James Sallee, Joseph Shapiro, Deirdre Sutula, Itai Trilnick, Sofia Villas-Boas, Reed Walker, and Catherine Wolfram for thoughtful comments and feedback. I also thank seminar participants at Case Western Reserve University, Energy Institute at Haas, Rutgers University, Santa Clara University, University of Alberta, University of California Berkeley, University of Connecticut, University of Maryland College Park, University of Massachusetts Amherst, University of Nevada Reno, University of Utah, US Department of Justice, AERE Summer Conference, Heartland Environmental and Resource Economics Workshop, and Columbia Interdisciplinary PhD Workshop in Sustainable Development.

market structure, empirical work on this topic has typically relied on assumptions about firm behavior or the nature of competition for identification.

In this paper, I use a novel source of exogenous variation in market size to estimate the causal effect of market size on market power in the Texas electricity market. Electricity is moved from producers to consumers over a grid of transmission lines, but the carrying capacity of transmission lines is limited. When this limit is hit, transmission congestion splits the statewide market into smaller "local markets," shrinking the set of rivals with which each power plant competes. For example, in the afternoon, a transmission constraint binds and a corner of the grid splits off from the full market into its own local market with fewer competitors; by that evening, the congestion subsides and the entire market is once again full integrated. Theory tells us that, for a power power plant in this corner of the grid, the incentive and ability to exercise market power vary greatly between these two times — when competing with fewer plants to supply electricity, a profit-maximizing firm will exercise market power by increasing markups. To test this theory, I exploit exogenous but predictable changes in market size due to transmission congestion and estimate how they affect noncompetitive firm behavior.

Electricity markets provide a useful context to study market power because they simplify many of the common challenges to identification and estimation. First, electricity is a homogeneous good, which removes issues of product differentiation. Second, demand is perfectly inelastic and observed, so demand estimation is not required. Third, power plants have simple production functions with marginal costs that are easily calculated using available data. Fourth, deregulated electricity markets clear through a centralized auction mechanism with specified rules, so the game being played is common to all firms and known to the researcher. For these reasons, many prior studies of market power have focused on electricity markets,[2] which I build on for this work.

I use detailed microdata on power plant-specific pricing behavior and local market structure, which are available in the data-rich Texas electricity market. I observe every power plant's offer curve — a function that indicates how much the plant will produce at each price — at 15-minute intervals. I also have data on power plant operations and fuel prices, which I use to calculate every plant's marginal cost. I combine these offer curves and cost data to calculate plant-specific markups every 15 minutes. I also observe the presence of transmission congestion and, hence, local market size at 5-minute intervals. Importantly, these detailed microdata on markups and local market size exhibit variation both cross-sectionally and over time at a high frequency.

In order to estimate the causal effect of local market size on markups, I exploit exogenous variation in transmission congestion due to temperature shocks. Transmission capacity is constrained by thermal limits on transmission lines, so less electricity can be transmitted at higher ambient temperatures. Hot temperatures also increase the consumption of electricity in Texas due to demand for indoor cooling. Thus, at high temperatures, more electricity is demanded, but the electrical grid has less capacity to transmit, resulting in congestion and

[2]See Green and Newbery (1992), Borenstein and Bushnell (1999), Wolfram (1999), Borenstein, Bushnell, and Wolak (2002), Hortaçsu and Puller (2008), and Bushnell, Mansur, and Saravia (2008), among many others.

local markets. This congestion, however, may extend well beyond the hottest areas of the grid. Due to the complex structure of the network, a temperature shock at any point in the grid can alter transmission throughout the entire network, resulting in local markets to form in geographically distant parts of the grid. I leverage this fact in my identification strategy.

I use an instrumental variables approach, instrumenting for a power plant's local market size with temperatures throughout the transmission grid. Because temperature may affect many aspects of an electricity market, including the level of demand, I control flexibly for the local temperature at a power plant to account for all characteristics of the plant's local market that vary with temperature. I further control for unobservables through high-dimensional fixed effects, including an hour of sample fixed effect that accounts for time-varying shocks common to all plants in the market. This empirical strategy yields a relatively weak assumption to identify the causal effect of market size on markups — after controlling for both the local temperature near the plant and average temperatures through fixed effects, the remaining residual variation in geographically distant temperatures affects markups only through market size and through no other channels.

To select from many possible instruments — hourly temperature at 75 weather stations throughout Texas — for each power plant, I use machine learning techniques to create a plant-specific prediction model that forecasts each power plant's local market size. I find that, on average, a plant's local market size is best predicted by a nonparametric function of temperature at a single weather station, rather than a parametric function of multiple temperatures. Thus, I match each power plant to the weather station that best predicts its local market size and instrument with a nonparametric function of the matched temperature. This procedure is similar to the instrument selection method of Belloni et al. (2012). This prediction exercise also mimics the forecast that firms must make when submitting offer curves in advance of the market clearing; I show that roughly 45% of variation in local market size can be explained by a simple prediction model using data available to all firms when they submit offers.

I next show how local market size and markups vary with matched temperature. In the first stage, I find a power plant's local market shrinks as the temperature at the plant's matched weather station exceeds 30–35°C (86–95°F); at the extreme, a plant's local market is 9–12% smaller on average when the matched temperature exceeds 40°C (104°F), compared to moderate temperatures. The power plant's markups increase over the same range of matched temperatures; at temperatures over 40°C (104°F), markups are on average 135–195% larger than at moderate temperatures.

I then use two-stage least squares to estimate the causal effect of market size on markups. I find that when a power plant's local market shrinks by 10% of the mean market size, the firm increases markups at the plant by 120–160% of the mean value, or by more than 3.5 times the average marginal cost of a fossil fuel-fired power plant. This result is consistent with a model of firm behavior in which the firm sets markups at a power plant in response to residual demand at the plant. When the plant's local market is smaller, it faces residual demand that is less elastic, so the firm exercises market power by increasing markups. The magnitude of this result is large, showing that firms in this market respond to even a relatively small change in market size by substantially increasing the exercise of market power.

I finally calculate the welfare effects of this source of market power. To do so, I consider a counterfactual in which firms do not increase markups in response to transmission congestion induced by high temperatures. I assume these markups are passed through into higher average retail electricity prices in equilibrium, which has two main effects. First, although demand is inelastic in the real-time wholesale market, it has some elasticity to end-use consumption in the long run; higher retail prices reduce consumption in all hours, which generates \$7.1–21.5 million of deadweight loss annually. Second, consumers pay an extra \$2.1 billion per year for the electricity they consume, which is purely a transfer to producers and raises important equity concerns. I further calculate how these effects will change as a result of climate change. I find that warming of 2.5°C (4.5°F) will roughly double the welfare effects of markups induced by transmission congestion at high temperatures. These results show that long-run planning and investment in this sector would benefit greatly by considering the competitiveness effects of climate change.

This paper makes four main contributions to the economics literature. First, these results contribute to a literature on the relationship between market structure and competition. Empirical studies on this topic date back to the work of Bain (1951), who estimated the correlation between profits and industry concentration in U.S. manufacturing; Bresnahan (1989) provides a summary of additional work in this "structure-conduct-performance paradigm." More recent empirical work uses models of firm entry and exit to estimate the competitiveness of industries (Bresnahan and Reiss 1991). These studies, however, typically require strong assumptions, such as how firms compete or how the number of firms in the market affects firm profitability (Berry and Reiss 2007).

In this study, by contrast, I exploit an exogenous source of market structure variation, from which I estimate the causal effect of market size on market power using weaker assumptions than these previous approaches. Bergquist (2017) provides experimental evidence of this causal effect by incentivizing the entry of intermediaries in Kenyan agricultural markets. She finds no effects on market outcomes and infers that the entrants easily join the collusive agreements that existed among incumbents. These results differ greatly from my estimates and suggest that this effect varies by settings. My work is the first to estimate this effect under weak assumptions in an industrialized setting.

Second, this paper contributes to a literature on market power and strategic firm behavior in wholesale electricity markets. Green and Newbery (1992), Borenstein and Bushnell (1999), Wolfram (1999), Borenstein, Bushnell, and Wolak (2002), Hortaçsu and Puller (2008), and Bushnell, Mansur, and Saravia (2008), among many others,[3] document the presence of market power in wholesale electricity markets, quantify the costs of this market power, and characterize the type of competition that exists in the market. This is one of the first papers, along with McDermott (2018), to document how market size or market structure affects noncompetitive behavior in these markets. This is also one of the first papers, along with Mercadal (2018), to emphasize the unique way in which local markets and the resulting market power form in nodal electricity markets.[4]

[3]See Wolfram (1998), Puller (2007), Fabra and Reguant (2014), Reguant (2014), Ito and Reguant (2016), and Hortaçsu et al. (2017) for more examples in this literature.

[4]The Texas electricity market uses nodal pricing, meaning prices are determined at every node in the network to reflect the marginal value of electricity at that point in the grid. With no binding constraints,

Third, this work contributes to a literature on the role and value of transmission infrastructure in wholesale electricity markets. Borenstein, Bushnell, and Stoft (2000) and Joskow and Tirole (2000) provide theoretical support for the competitiveness effects of transmission infrastructure. In empirical work, Davis and Hausman (2016) exploit the closure of a nuclear power plant to value inter-regional transmission and document suggestive evidence of market power during hours of transmission congestion. Most closely related to this work, Ryan (2017) uses structural estimation to simulate an expansion of inter-regional transmission in the Indian electricity market. This is the first paper to directly estimate the effect of transmission congestion, and the resulting local markets, on firm behavior and the exercise of market power.

Finally, this paper contributes to the growing literature that examines the effects of climate on many economic outcomes, including effects on energy consumption (Auffhammer and Mansur 2014; Auffhammer, Baylis, and Hausman 2017; Auffhammer 2018), agriculture (Schlenker, Hanemann, and Fisher 2005; Deschênes and Greenstone 2007; Schlenker and Roberts 2009), productivity and growth (Dell, Jones, and Olken 2012; Burke, Hsiang, and Miguel 2015), conflict (Hsiang, Meng, and Cane 2011; Hsiang, Burke, and Miguel 2013), and mortality (Deschênes and Greenstone 2011). This is the first study to find climate change effects on market structure and competition, as well as the first to find effects on the supply of electricity.

1.2 Texas electricity market

The wholesale electricity market in Texas is managed by the Electricity Reliability Council of Texas (ERCOT). ERCOT is the independent system operator (ISO) that schedules power plants to supply electricity and coordinates the use of the transmission grid to deliver that electricity to consumers. Although representing only one state,[5] this market includes more than 610 individual generating units and roughly 25 million end-use consumers, with more than 46,500 miles of high-voltage transmission lines connecting producers and consumers (ERCOT 2018). In the years 2011–2014, nearly $12 billion was transacted annually in ERCOT's real-time wholesale market.

ERCOT is tasked with maintaining a reliable supply of electricity to consumers throughout its territory. Unlike the markets for other commodities, the wholesale market for electricity must be operated by an ISO due to the unique nature of electricity. First, storage of electricity is prohibitively expensive, so production of electricity must perfectly meet demand in real time to ensure an uninterrupted supply. Second, transmission of electricity is subject to the physical constraints of the electrical grid, such as thermal limits that constrain the quantity of electricity that can be transmitted. Hence, production of electricity must

prices are uniform at all power plants. When transmission constraints bind, nodal prices diverge to reflect the shadow value of relevant constraints. This form of pricing is in contrast to zonal pricing, in which the market is divided into several zones and only inter-zonal transmission constraints are considered when setting market-clearing prices. Thus, a nodal market has orders of magnitude more possible configurations of local markets than does a comparable zonal market.

[5]The ERCOT territory covers roughly 75% of the geographic area of the state of Texas, but this area accounts for approximately 90% of the electricity consumed in the state.

also occur at the correct locations to ensure the transmission grid is not overloaded and compromised.

A second mandate of ERCOT is to provide this reliable supply of electricity at lowest cost. To achieve this goal, ERCOT uses an auction mechanism in which each firm that produces electricity submits an offer curve — a function that indicates how much the plant will produce at each price — for each power plant it owns. ERCOT awards production to the power plants with the lowest-price offers, subject to the constraints described above. When any of these constraints bind, the marginal cost of electricity can vary across the electrical grid. ERCOT calculates this marginal cost at more than 1000 nodes throughout the grid and assigns each a locational marginal price (LMP) that reflects the shadow value of relevant constraints.

The Texas electricity market is unique among electricity markets in the United States for two main reasons, both of which make this market an attractive setting for my study. First, the ERCOT grid is its own interconnection, meaning there is little capacity for trade with adjacent electricity markets. Thus, I do not consider imports or exports in this study.[6] Second, ERCOT operates an energy-only market[7] with relatively loose limits on prices[8] and firm behavior.[9] These factors reduce the possibility that my estimates are attenuated due to overlapping markets or regulatory restrictions.

1.2.1 Real-time market

The focus of this study is ERCOT's real-time wholesale electricity market.[10] This market clears every five minutes and determines which power plants produce electricity and how transmission is managed to ensure reliable delivery of electricity to end-use consumers. If a producer wants to generate electricity, it must participate in the real-time market by submitting an offer curve.[11] This curve specifies how much electricity a power plant will produce at every price, and the firm submits a separate curve for each power plant it owns.

[6] By contrast, an important component of the Borenstein, Bushnell, and Wolak (2002) study of the California electricity market is the import of electricity into the state because California is one of several markets within the larger Western Interconnection.

[7] Many electricity markets are accompanied by a capacity market that pays owners of power plants for their capacity, in addition to actual production, to ensure the market has adequate generation resources.

[8] ERCOT began this study period with a price cap of $3,000 per MWh, which it increased to $9,000 per MWh by the end of this study period. Many other electricity markets, by contrast, have price caps of $1,000 per MWh.

[9] Most electricity markets have procedures to mitigate excessively high offer prices. ERCOT does so only in the most extreme cases, which I describe in Appendix A.1.

[10] Firms offering to produce electricity in the real-time market may have existing financial positions due to long-term bilateral contracts or ERCOT's day-ahead market. These positions are purely financial, however; physical delivery of electricity is determined only in the real-time market. See Appendix A.1 for details on these other markets.

[11] Alternatively, a firm can set an output schedule for a power plant, which specifies a quantity to produce; ERCOT assigns the power plant to produce the specified quantity regardless of the market-clearing price. This option is used infrequently.

An offer curve must be weakly increasing in price and piecewise linear with up to 35 vertices.[12] The top panel of Figure 1.1 shows an example offer curve. I observe every offer curve submitted to ERCOT, which allows me to directly observe pricing behavior in this market without making additional assumptions about firm behavior.

On the demand side of the real-time market, purchasers of wholesale electricity are primarily retail providers that purchase electricity to distribute to their end-use consumers. These retail providers contract with their customers to provide electricity at retail rates that rarely vary with the wholesale price in the real-time market.[13] As a result, demand is considered to be inelastic in real time and the market clears on the supply side; ERCOT dispatches power plants to meet demand at lowest cost, subject to the physical constraints of the transmission grid. The resulting prices are equal to the marginal cost of electricity production at each node of the grid, as discussed below. The bottom panel of Figure 1.1 depicts the simplest example, when no constraints bind; the full market clears where the aggregate offer curve intersects the vertical demand curve, and all producers receive the market-clearing price of roughly $40 per MWh.

A firm offering to produce electricity in this market has great flexibility in its strategy throughout the course of a day; a power plant may have a different offer curve for each hour of the day. Additionally, a firm may adjust its offer curves up to the beginning of the preceding hour,[14] which allows the firm to refine its strategy as new information becomes available as the market time approaches. Much of the information a firm may consider when submitting its offer curves is known to all participants or is forecastable over this time frame. For example, firms have knowledge of the marginal cost of all power plants,[15] as well as which plants are offline at any point in time. Demand is highly forecastable, particularly within an hour or two; in fact, ERCOT publishes hourly demand forecasts beginning one week in advance. Transmission congestion and the formation of local markets are also forecastable, which I discuss next and model in Section 1.5.

1.2.2 Transmission congestion and local markets

An important component of ensuring a reliable supply of electricity is ERCOT's management of the transmission grid. Transmission lines move electricity around the ERCOT region, but thermal constraints limit the maximum carrying capacity of these lines. Transmission of electricity generates heat in the lines, and the amount of heat produced is proportional to the quantity of electricity being transmitted. When transmission lines overheat they can

[12] Firms submit the price-quantity pairs that define the vertices of offer curves, and the offer curve is linearly interpolated from these points. Each curve can have up to 35 vertices, but, in practice, firms often use many fewer points when specifying offer curves.

[13] Retail providers may offer time-of-use plans with rates that vary over the course of a day and seasonally. Although these plans are typically designed to follow the general price patterns observed in the real-time wholesale market, the rates paid by the consumers are set when the contract is signed and do not respond in real time to the wholesale market.

[14] For example, offer curves for the 2–3 p.m. operating hour can be adjusted up to 1 p.m.

[15] The marginal cost of producing electricity at a given power plant is primarily the cost of fuel, or the product of the plant's heat rate — the rate at which the plant converts fuel into electricity — and the fuel price. Data on heat rates are publicly available, as are spot prices for common fuels.

fail, which not only renders that transmission line inoperable, but can also cause cascading failures throughout the electrical grid.[16] As a result, it is critically important that the ISO ensures the dispatch of power plants, and the subsequent flow of electricity to consumers, does not cause a transmission line to exceed its operable limit.

To account for these constraints while still meeting electricity demand as efficiently as possible, ERCOT implements locational marginal pricing. That is, each location, or node, throughout the electrical grid can have its own price that equals the marginal cost of electricity at that point in the network. For example, when the lowest-cost dispatch of power plants does not violate any constraints, the market clears as a uniform-price auction and all LMPs are equal. However, if this dispatch would force a transmission line to exceed its carrying capacity, ERCOT must deviate from the lowest-cost dispatch, and LMPs diverge as some LMPs incorporate the shadow value of the binding constraints. Figure 1.2 displays each of these scenarios. In the top panel, no constraints bind and all producers receive the market-clearing price of roughly $40 per MWh. In the bottom panel, however, a constraint limits the flow of electricity into the southern portion of the grid, causing the LMPs in this region increase substantially, to more than $3,500 per MWh.

When transmission lines become congested, the ERCOT region is no longer fully integrated, as indicated by the divergence of LMPs. Any additional demand in the higher-priced region cannot be met with additional production in the lower-priced region but instead must be met locally. Thus, these power plants in the higher-priced region do not compete on the margin with power plants in the lower-price region, only with other plants in the same region. I refer to each of these regions as a "local market." In other words, the local market is the set of power plants within the ERCOT region at large that share a common LMP and compete to supply electricity within that region. When no transmission lines are congested, every power plant's local market is the entire ERCOT market, as depicted in the fully integrated market in the top panel of Figure 1.2. At the other extreme, when transmission lines are heavily congested, a very small local market forms and its LMP differs greatly from nearby local markets, as in the bottom panel of Figure 1.2.

By changing the number and kinds of power plants that compete with one another, the formation of local markets changes the ability of any one firm to exercise market power within a local market. In particular, when a local market contains fewer firms or less production capacity, economic theory suggests that market will be subject to greater market power, as I show in Section 1.3. With fewer power plants in its local market, the residual demand facing any one plant is less elastic, providing greater opportunity to exercise market power and influence the resulting LMP. I exploit this variation in local markets to estimate the effect of market size on markups.

Because the physical limits on transmission are thermal in nature, the carrying capacity of a transmission line is affected by the ambient air temperature; at high temperatures, a transmission line can carry less electricity while staying within the operable thermal limits.

[16] Transmission lines experience thermal expansion, causing the lines to lengthen and sag as the internal temperature increases. If lines expand excessively, they can contact other transmission equipment or nearby trees, causing the system to short circuit or, in an extreme case, create sparks that could start a fire. Other components of the transmission system are also put under greater stress at higher temperature, increasing the incidence of component failure in both the short run and long run.

For example, an increase from 38°C to 43°C (100°F to 109°F) reduces transmission capacity by 7-8% (Sathaye et al. 2013). Additionally, a substantial portion of electricity demand in the ERCOT region is for indoor cooling and, to a lesser extent, heating, so the ambient temperature drives the quantity of electricity being transmitted throughout the electrical grid. At high temperatures, more electricity is consumed, but the electrical grid has less capacity to transmit electricity, resulting in higher rates of transmission congestion and the formation of smaller local markets. In Section 1.5 I show that the formation of local markets is forecastable and develop a predictive model of each power plant's local market as a function of ambient temperature. This model suggests a firm is able to forecast the local market for each of its power plants and adjust their offer curves accordingly. I exploit the exogenous and quasi-experimental variation in local markets that is induced by temperature in order to causally identify the effect of market size on markups.

1.2.3 Power plants and firms

The Texas electricity market consists of 73 firms that operate 390 power plants[17] to supply electricity during my study period. Table 1.1 shows summary statistics for these plants by fuel type. Four types of plants — natural gas, coal, renewables, and nuclear — account for nearly all of the capacity and production in this market. Each type of power plant has unique technical characteristics that influence how it is offered to supply electricity and how it may affect the ability of other firms to exercise market power, which I describe in more detail in Sections 1.3 and 1.4. Figure 1.3 displays a map of units by fuel type, showing that plants are generally spread throughout the state, other than renewable plants, which cluster in areas with the greatest wind resources.

Table 1.1 also shows summary statistics by firm for the eight largest firms in this market. By many traditional metrics, this market does not appeared concentrated to the level that market power should be a concern. The three largest firms have markets shares of 20%, 15%, and 10%, respectively, when measured by production, and even less when measured by capacity; 60 firms have a market share of less than 1% by either metric. The Herfindahl–Hirschman Index for this industry is 0.059 when measured by capacity and 0.086 when measured by production, neither of which implies an uncompetitive market.[18] Traditional measures of competitiveness, however, are poorly suited for this industry because of the unique nature of electricity, including the lack of storage and the constrained nature of the grid on which the market operates.

[17] I use the terms "power plant" and "unit" interchangeably to describe the cross-sectional unit of my analysis, which I keep as disaggregated as possible while still allowing me to accurately join different data sources. In many cases, a single facility comprises two or more "units" by my definition. For example, if a coal-fired power plant has two separate boilers, each connected to its own generator, which can be operated independent of one another, and data are reported separately for each boiler-generator pair, then I consider this facility to be two separate units.

[18] The Herfindahl–Hirschman Index (HHI) is a measure of market concentration, calculated as the sum of squared market share. There is no universally agreed upon threshold to define an industry as competitive, but rules of thumbs would generally consider an industry with HHI less than 0.1 to be competitive.

1.3 Theoretical framework

To formalize the behavior of firms in ERCOT's real-time wholesale electricity market, I develop a model of oligopoly competition in this setting. This model closely follows that of Ryan (2017), who also models firm behavior under transmission constraints, but in the Indian electricity market. I adapt some notation from Hortaçsu and Puller (2008), who study a previous iteration of the Texas electricity market. Both of these previous models, as well as the model I develop, are auctions for shares (Wilson 1979). I first generate an optimal markup rule and then show how optimal markups change as market size changes.

1.3.1 Optimal markup rule

Each firm submits an offer curve, $S_i(p)$, for each power plant it owns, where i denotes plant. This offer curve specifies how much the plant will produce at each price and is defined by a vector of offer prices, $\mathbf{b}_i$, and a vector of offer quantities, $\mathbf{q}_i$.[19] The cost to produce at the plant is $C_i(q)$. The profit earned at plant i is $\pi_i = S_i(p_i)p_i - C_i(S_i(p_i))$, where p_i is the LMP that results at plant i.

Firms submit offer curves in the face of several sources of uncertainty. First, the market may be fully integrated or transmission congestion may split the market into two or more local markets. Firms are able to partially forecast transmission congestion,[20] but not perfectly, so local markets are stochastic from the perspective of a firm submitting offer curves. I denote the set of local markets as $\mathcal{M}$. Second, demand is also highly predictable but ultimately stochastic. I denote demand in market m as D_m, which I assume is perfectly inelastic.[21] Third, firms do not know the offer curves of other firms. Because this is a repeated game, firms may have an expectation of other firms' offer curves. Firms also typically know which plants are online and ready to produce. They do not, however, have full information on other offer curves. A final consideration of a firm when setting offer curves is any forward position that the firm has.[22] I denote the price and quantity of firm f's forward position in market m as P^F_{fm} and Q^F_{fm}, respectively. Although a firm knows its forward positions, the distribution of those positions over local markets is uncertain because transmission congestion is stochastic.

Each firm maximizes its expected profit summed over all of its plants and existing forward positions:

$$\max_{\{S_i|i\in\mathcal{I}_f\}} \mathbb{E}\left[\sum_{i\in\mathcal{I}_f}(S_i(p_i)p_i - C_i(S_i(p_i))) + \sum_{m\in\mathcal{M}}(P^F_{fm} - p_m)Q^F_{fm}\right] \tag{1.1}$$

[19] I place no restrictions on offer curves in this model. ERCOT gives firms great flexibility in setting offer curves, as I describe in Section 1.2.

[20] See Section 1.5 for a prediction model of local market size.

[21] This assumption is common in the literature and reflects the reality of wholesale electricity demand. See Section 1.2 for a brief discussion of demand in ERCOT's wholesale electricity market.

[22] See Appendix A.1 for a discussion of bilateral contracts and the day-ahead market. I consider these positions to be exogenous from the real-time market.

where $\mathcal{I}_f$ is the set of plants owned by firm f and p_m is the common LMP that resolves in market m. The final term is the profit or loss from forward positions, which the firm effectively must unwind in the real-time market.

The market clears by setting the price in each local market such that production equals demand net of flows between markets:

$$\sum_{i \in \mathcal{I}_m} S_i(p_m) = D_m + T_m$$

where $\mathcal{I}_m$ is the set of plants in market m and T_m is net transmission flows out of market m.[23] Rewriting this condition from the perspective of plant i:

$$S_i(p_m) = D_m + T_m - \sum_{\substack{j \in \mathcal{I}_m \\ j \neq i}} S_j(p_m) \tag{1.2}$$

Note that the right-hand side of this equation is the residual demand of plant i, which I denote $R_i(p_m)$.

To get the firm's first-order condition, I make two assumptions. First, every plant's marginal cost is constant in quantity up to the plant's capacity constraint. I denote plant i's marginal cost as c_i. Second, I assume no constraints are marginally binding. This assumption implies that, on the margin, offer price does not affect whether a constraint binds. I also define Q_{fm} to be the production by firm f in market m net of the firm's forward position in the market. That is, $Q_{fm}(p_m) = \sum_{i \in \mathcal{I}_f \cap \mathcal{I}_m} S_i(p_m) - Q^F_{fm}$.

Taking the derivative of the firm's objective function, from Equation (1.1), with respect to plant i's kth offer price, b_{ik}, and substituting in the residual demand, from Equation (1.2), yields the first-order condition:[24]

$$b_{ik} - c_i = \frac{\mathbb{E}[Q_{fm}(b_{ik})]}{-\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]} \tag{1.3}$$

where firm f is the firm that owns plant i. This result is the optimal markup rule for plant i's kth offer quantity, similar to the Lerner index. The numerator is the firm's expected production that will be paid market m's real-time price when this offer price clears the market; when a firm has more inframarginal production in a market, it has a greater incentive to exercise market power. The denominator is the negative of the slope of the plant's expected residual demand when this offer price clears the market; when residual demand is less elastic, a firm has a greater ability to exercise market power.

1.3.2 Comparative statics

From the optimal markup rule in Equation (1.3), I derive how markups change in response to changes in the market size, which I then estimate in Section 1.6. I denote the local market

[23] Electricity only flows from lower-priced markets to higher-priced markets, so T_m does depend on prices. However, local markets are defined by binding constraints, so marginal changes in a market's price will not change flows of electricity between local markets. Thus, I consider T_m to be inelastic on the margin.

[24] See Appendix A.2 for a full derivation of this first-order condition.

size for plant i as K_i, which I measure as the total production capacity in the local market. In my empirical analysis, I consider two additional measures: the number of units in the local market and the number of firms in the local market; theoretical results are comparable for all three measures.

The percent change in markup, $b_{ik} - c_i$, due to a change in the local market size is:[25]

$$\frac{d(b_{ik} - c_i)/dK_i}{b_{ik} - c_i} = \frac{\partial \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i}{\mathbb{E}[Q_{fm}(b_{ik})]} - \frac{\partial \mathbb{E}[\partial R_i(b_{ik})/\partial p_m]/\partial K_i}{\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]} \tag{1.4}$$

The first term is the percent change in the optimal markup rule's numerator, and the second term is the percent change in the optimal markup rule's denominator. Thus, the direction of the effect depends on whether the firm expects a change in market size to have a greater percent effect on its own production in the local market — more production gives the firm a stronger incentive to exercise market power — or on the slope of expected residual demand — less elastic residual demand gives the firm a greater ability to exercise market power. If the first term dominates, then a reduction in the local market size would yield smaller markups. If the second term dominates, then a reduction in the the local market size would cause markups to increase.

In theory, this effect could go in either direction. In practice, however, the second term is likely to dominate in my empirical setting. As I show in Section 1.4, transmission congestion occurs much more often when the ambient temperature is high. Hot temperatures also result in a higher level of demand, meaning the market clears further out on the aggregate offer curve. This outer portion of the aggregate offer curve tends to become very inelastic. When the market is operating under these conditions, even small changes in local market size may cause large changes in the slope of residual demand. Thus, I expect the second term to dominate, resulting in a negative effect — firms set larger markups when the local market size is reduced.

I further examine how this effect varies depending on the types of capacity in the local market. I first consider how a plant's markups are differently effected by a change in local market capacity owned by competitors, K_i^{comp}, and a change in local market capacity owned by its own firm, K_i^{own}. Ownership has an important effect on the first term of the comparative static in Equation (1.4):

$$\frac{\partial \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i^{comp}}{\mathbb{E}[Q_{fm}(b_{ik})]} < \frac{\partial \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i}{\mathbb{E}[Q_{fm}(b_{ik})]} < \frac{\partial \mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i^{own}}{\mathbb{E}[Q_{fm}(b_{ik})]}$$

Changes in the capacity of competitors have no direct effect on a firm's own production, whereas changes in its own capacity have large effects. Thus, the first term dominates most often when considering a firm's own capacity and least often when considering competitors:

$$\frac{d(b_{ik} - c_i)/dK_i^{comp}}{b_{ik} - c_i} < \frac{d(b_{ik} - c_i)/dK_i}{b_{ik} - c_i} < \frac{d(b_{ik} - c_i)/dK_i^{own}}{b_{ik} - c_i} \tag{1.5}$$

In words, the effect of competitors on markups is more negative than the average effect. On the other hand, the effect of a firm's own capacity is less negative and may even change the sign of the effect — a firm may set lower markups when it has less capacity in a local market.

[25] See Appendix A.2 for a full derivation of comparative statics.

I finally examine how different types of plants affect markups. I consider renewables and nuclear plants to be nondispatchable because they almost always produce at maximum capacity; these plants have low marginal costs and typically submit offer curves that are almost perfectly elastic. I consider all other plants to be dispatchable because their production tends to follow the market-clearing price; these plants have higher marginal costs and typically submit offer curves that slope upward. Plant type has an important effect on the second term of the comparative static in Equation (1.4):

$$\frac{\partial \mathbb{E}[\partial R_i(b_{ik})/\partial p_m]/\partial K_i^{nd}}{\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]} < \frac{\partial \mathbb{E}[\partial R_i(b_{ik})/\partial p_m]/\partial K_i}{\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]} < \frac{\partial \mathbb{E}[\partial R_i(b_{ik})/\partial p_m]/\partial K_i^{disp}}{\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}$$

where K_i^{nd} is nondispatchable capacity and K_i^{disp} is dispatchable capacity. Changes in nondispatchable capacity only shift residual demand, whereas changes in dispatchable capacity rotate residual demand, due to the elasticity of offer curves submitted for each group. Thus, the second term dominates most often when considering dispatchable capacity and least often when considering nondispatchable capacity:

$$\frac{d(b_{ik} - c_i)/dK_i^{disp}}{b_{ik} - c_i} < \frac{d(b_{ik} - c_i)/dK_i}{b_{ik} - c_i} < \frac{d(b_{ik} - c_i)/dK_i^{nd}}{b_{ik} - c_i} \tag{1.6}$$

In words, the effect of dispatchable capacity on markups is more negative than the average effect, whereas the effect of nondispatchable capacity is less negative than average.

1.4 Data

To estimate the effect of market size on markups, I construct a novel panel dataset at the power plant-by-hour level for the years 2011–2014. This dataset has four important features: offer curves, marginal costs, local market sizes, and temperatures. I discuss each of these data sources and describe how I construct the main variables of interest.

1.4.1 Offer curves

I observe all offer curves that firms submit to ERCOT's real-time wholesale electricity market. These data are from ERCOT's daily 60-Day SCED Disclosure Reports.[26] Each report includes the sets of price-quantity pairs that trace out every unit's offer curve at 15-minute intervals.[27] From these offer curves I directly observe the pricing strategy of firms at every power plant, and, when combined with my calculation of marginal, I use them to calculate markups. Figure 1.4 shows several example offer curves, including one for each of the main fuel types.

[26]The 60-Day SCED Disclosure Reports are so-called because the release of each report is delayed 60 days to reduce the ability of firms to exercise market power by best-responding to observed offers of competitors.

[27]The 60-Day SCED Disclosure Reports also include the operating status of each unit at 15-minute intervals. I omit a unit when its status indicates it is not being dispatched by the typical rules of the ISO, such as when a unit is down for maintenance or undergoing testing procedures.

The natural gas and coal units share a similar shape. At low quantities, these units are offered at low or negative prices for dynamic reasons; many of these power plants are costly to shut down and start up again, so a firm is willing to accept a price below marginal cost for a short period of time to avoid incurring these fixed costs. The middle range of quantities is offered at a price roughly equal to marginal cost. At higher quantities, however, offer prices are above marginal cost and increase with quantity in a way that is not consistent with marginal cost pricing.

The wind and nuclear units are each offered at negative prices, but for different reasons. The wind unit has no fuel cost, so its marginal cost is approximately zero. It does, however, generate a federal production tax credit when it produces electricity. Thus, the opportunity cost of this unit is not zero but rather the negative of the tax credit, and the unit's entire capacity is offered at this price. Nuclear units, on the other hand, are designed to operate at full capacity, and deviating from full capacity is costly. Thus, the owner of a nuclear unit will accept negative prices to ensure the unit continues to produce at full capacity.

The final two example offer curves in Figure 1.4 show the same natural gas unit during two different hours. In both hours, the unit has a marginal cost of roughly $35 per MWh. The offer prices at higher quantities, however, exceed $150 per MWh in one example and $800 per MWh in the other. These offer curves are not consistent with the marginal cost of the unit. They are, however, consistent with the exercise of market power by marking up prices well above marginal cost, which is the behavior I seek to identify in this study.

1.4.2 Marginal cost

The primary component of a power plant's variable cost is the cost of fuel.[28] I calculate each unit's variable fuel cost as the product of the unit's average heat rate — the rate at which the unit converts fuel into electricity — and the fuel price. Additionally, many fossil fuel-fired units are subject to emissions regulations and must purchase permits to cover their level of emissions. For these units I also calculate the variable cost of emissions, which I calculate as the product of the unit's average emissions rate and the permit price.[29] I sum these two variable costs to calculate marginal cost. I assume units have constant marginal cost up to the maximum capacity of the unit, as is common in the literature.

Data on fuel input and emissions are from the Environmental Protection Agency's (EPA) Continuous Emissions Monitoring System (CEMS), and data on net generation[30] are from ERCOT's 60-Day SCED Disclosure Reports. I combine these data to calculate average heat rates and average emissions rates.[31] Fuel prices are from several sources: S&P Global

[28] All power plants have a small variable cost of operations and maintenance. These costs are unobserved so I do not include them in my marginal cost calculation. In my empirical specification, unit-level fixed effects will account for these costs.

[29] Units face regulations for emissions of SO_2, NO_x, or both. I perform this calculation separately for each pollutant and sum over the pollutants for which the plant is subject to regulation.

[30] Net generation is the amount of electricity that is output to the grid. This measure is in contrast to gross generation, which is the total amount of electricity produced, some of which is consumed within the power plant itself.

[31] See Appendix A.3 for a more detailed discussion of these calculations, including how I impute missing data.

Market Intelligence, the Energy Information Administration's (EIA) Form 923, and Deutch et al. (2009); natural gas prices vary daily and regionally, coal prices vary monthly, and nuclear prices vary annually. Emissions prices are from S&P Global Market Intelligence and the EPA's market progress reports; emissions prices vary daily or monthly. As described above, I combine these prices with heat rates and emissions rates to calculate each unit's constant marginal cost.

With data on both offer prices and marginal cost, I calculate markup as the difference between price and cost. In most settings, markup is simply a single value. In this market, however, firms submit offer prices for every level of production, so markup is a function of the level of production. To normalize production quantities across power plants of different sizes, I use capacity factor — the plant's production as a percent of maximum possible production — as the measure of production. Figure 1.5 shows the average offer curve and average marginal cost, as a function of capacity factor, for all units and hours in my dataset, as well as the averages for each of the four primary types of units.[32]

In order to analyze markups, which are functions rather than single values, I select several points on each unit's offer curve at which I calculate the markup and estimate the effect of market size: 85%, 75%, and 65% of nameplate capacity. A power plant typically cannot offer 100% of its nameplate capacity to the market because some of its production is consumed within the plant, so it is common to observe a plant offer only 90% of its capacity.[33] Additionally, operating near full capacity can reduce the efficiency of the unit[34] and increase the need for maintenance in the long run. Production above 85% of capacity may be subject to one of both of these issues, so I do not consider markups at these high levels of production. On the other hand, below 65% of nameplate capacity, coal-fired units have offer prices below marginal cost on average, due to the plant dynamics I describe above. Thus, markups at lower levels of production are likely to be driven by these operational considerations and not the potential to exercise market power.

1.4.3 Local market size

I infer local market size from the price data in ERCOT's real-time price reports. These reports give LMPs at every node in the transmission network for every clearance of the market, which occurs at five-minute intervals. As an example, these are the data that I use to create Figure 1.2. As discussed in Section 1.2, any binding constraints cause a discrete change in LMPs equal to the shadow value of the constraint. I consider the transmission grid between two power plants to be unconstrained if the LMPs at those plants are within

[32]Each power plant's offer curve must be weakly increasing, but a firm can offer less than the full capacity at any of its plants. In fact, most plants cannot offer 100% of their capacity to the market because some production is consumed at the plant. As a result, the sample changes slightly as quantity increases, particularly as quantity exceeds 90%. For this reason, offer curves may appear to slope downward on average even though each unit-level curve is weakly increasing.

[33]The amount of production consumed at the plant is often at least 5% and typically closer to 10%. For example, Puller (2007) assumes a plant is operating at "full capacity" when it produces at 90% of nameplate capacity.

[34]Bushnell and Wolfram (2005) show that heat rate increases when a unit exceeds the 98th percentile of its output but is relatively constant at more moderate levels of production.

$1 per MWh or 1%, whichever is greater.[35] I then define a local market to be the set of all plants that can be connected by unconstrained transmission.[36] For example, in the bottom panel of Figure 1.2, I define one local market to be the lighter points in the southern portion of the grid, and I define a second local market to be the darker points throughout the rest of the grid. With local markets defined, I sum the production capacity, count the number of units, and count the number of firms within a local market to construct three measures of its size.

1.4.4 Temperature

Temperature data are from the National Oceanic and Atmospheric Administration's (NOAA) Integrated Surface Database. This database includes hourly — and occasionally more frequent — weather recordings from 35,000 weather stations worldwide. I limit my sample to weather stations in the state of Texas with minimal missing data.[37]

As described in Section 1.2, temperature is an important driver of the formation of local markets. Figure 1.6 depicts this relationship. The top panel shows the fraction of time that the market is either fully integrated or split into local markets as a function of the mean temperature in the state.[38] At most temperatures, the market is integrated a majority of the time. At mean temperatures above 30°C (86°F), however, the formation of local markets is much more prevalent, occurring roughly 75% of the time. The bottom panel of Figure 1.6 shows the frequency with which the market splits into a certain number of local markets. In particular, it is more common for the market to split into seven or more local markets when the state experiences high temperatures. Thus, when the mean temperature is greater than 30°C (86°F), not only are local markets more likely to form, but they are typically more numerous and, hence, smaller in size.

1.4.5 Additional data

I supplement these data with several additional data sources. Unit characteristics, such as fuel type, capacity, and location, are from the EIA's Form 860. Ownership data are from S&P Global Market Intelligence. System-wide demand data, which I use in my prediction models, are from ERCOT's 48-Hour Aggregate Load Summary.

[35] Due to the complex nature of this network, LMPs sometimes differ by a few cents, and it is not clear if these small price differences represent real constraints of the system. By imposing the requirement that LMPs must exceed $1 per MWh and 1%, I construct a conservative measure of local market size.

[36] For example, suppose LMPs are $30.00, $30.50, $31.25 at plants A, B, and C, respectively. I consider A and B to be connected by unconstrained transmission, and I also consider B and C to be connected by unconstrained transmission. Because A and C can be connected through B, I define all three plants to share a local market, even though A and C have LMPs that differ by more than $1.00.

[37] A weather station must have data for at least 97.5% of the hours during my study period and no more than 60 consecutive hours of missing data. See Appendix A.3 for a description of how I impute missing temperatures for the weather stations in my sample.

[38] I calculate mean temperature as the unweighted mean of all 75 weather stations in the state.

1.5 Empirical strategy

The goal of this analysis is to examine how markets size affects competition. The most straight-forward method to estimate the role of market size on market competitiveness is to simply regress a measure of competition on a measure of market size. For example, in this analysis, that would involve regressing markups on one of my measures of market size: market capacity, number of units in the market, or number of firms in the market. This method, however, has several empirical problems, both in general and because of this specific context.

First, in the general case, both market competitiveness and market size are equilibrium outcomes that are co-determined and depend importantly on underlying market characteristics. In the absence of exogenous variation in market size, any attempt to causally estimate this effect will suffer from simultaneity bias. Second, in this specific case, the complex nature of the transmission grid means that markups can effect the definition of a market. Offer curves and markups determine where electricity is produced and how it is transmitted to consumers, which may result in transmission congestion and the formation of local markets.

This setting features another potential problem; a firm must submit its offer curves to the system operator one to two hours before the market clears, when the firm does not have perfect information about the market in which it will operate. If the local market size was determined entirely at random after firms had submitted offer curves, then markups would be uncorrelated with the ultimate market size. In this case, the simple regression of markups on market size would yield no significant effect, even if a firm would have exercised market power had it known the size of the market in which it would be operating.

To overcome the potential endogeneity of market size and to isolate predictable variation in local market size, I use an instrumental variables approach, instrumenting for market size with contemporaneous temperature. In order for temperature to be a valid instrument, however, temperature must be uncorrelated with markups other than through its effect on local market size. If this is not true, then the estimated effect of local market size on markups is confounded with any other channels through which temperature may influence the markup decisions of firms.

The primary concern in this setting is that the temperature near a power plant may affect the characteristics of that plant's local market through more channels than just the size of the market. In particular, the level of demand in this market is strongly driven by temperature due to indoor cooling and heating. If an hour of extreme heat both induces transmission congestion and increases demand, then local markets are smaller and clear further out on the aggregate offer curve, corresponding to a less elastic supply of electricity. Both effects decrease the elasticity of residual demand faced by a plant in this local market, which increases the incentive to exercise market power. But an empirical strategy that instruments with the local temperature would attribute this entire effect to the size of the local market and ignore the effect of other changes in local market conditions.

To overcome this threat to identification, I do not use the local temperature at each power plant as an instrument. Instead, I use temperatures at all 75 weather stations throughout the state of Texas as potential instruments. Due to the complex nature of the transmission grid, a temperature shock at any point in the network can affect the operation of the entire

grid, including the incidence of transmission congestion and the formation of local markets. Because these other temperatures are geographically farther from the power plant, they are less likely to be correlated with market conditions in the plant's local market. Additionally, I can control for the local temperature closest to the power plant to isolate temperature variation that is orthogonal to any temperature-related characteristics of that plant's local market.[39]

To determine which of the 75 weather stations to use as instruments and what functional form these temperatures should take, I develop a prediction model of local market size. I next discuss this prediction model and its results, and I then present my empirical specification that incorporates the results of the prediction model.

1.5.1 Prediction model

Ambient air temperature is a key driver of local market formation, as described in Section 1.2 and depicted in Figure 1.6. I use machine learning tools to develop a prediction model of local market size as a function of temperatures throughout Texas, as well as other explanatory variables. The results of this model determine the first stage of my two-stage least square empirical strategy. This methodology is similar to that of Belloni et al. (2012), but I consider additional variable selection procedures beyond the LASSO model of their approach.

In addition to determining the first stage of my two-stage least squares (2SLS) regression, this prediction modeling exercise mimics the prediction problems that firms face when submitting offer curves to the system operator. Because firms must submit their offers one to two hours before the market clears, I seek to develop a simple model using data that are available to firms when they submit offer curves. By doing so, my first stage isolates variation in local market size that is not only exogenous but also predictable by firms operating in this market.[40]

I predict hourly local market capacity for each power plant as a function of 75 temperatures throughout Texas, indicators for each hour of day within a month of the year, system-wide demand, and local market capacity lagged two hours. I consider five different model specifications, which I describe below, in order to explore the trade-off between estimating the effect of temperature on local market capacity as flexibly as possible for one weather station and including interactions between temperatures at multiple weather sta-

[39]This identification strategy is similar to that of Schlenker and Walker (2016). They study the effect of air pollution due to taxiing airplanes at California airports. The endogenous variable is the taxiing time of these airplanes, and they instrument with taxiing time at large airports in the Eastern United States, an indicator of flight delays that propagate throughout the flight network.

[40]With more than $12 billion transacted annually in ERCOT's real-time wholesale electricity market, firms benefit greatly from predicting markets and submitting optimal offer curves. The most sophisticated firms in this market have teams of employees to model future market outcomes. Consulting firms offer these services to firms that do not conduct forecasts in-house. I do not seek to duplicate these efforts. Instead, my prediction model mimics a forecast that could be conducted by even the less sophisticated firms in this market.

tions. In each case, I use stratified five-fold cross-validation to ensure the models are not overfitted.[41]

The first model simply accounts for the mean local market capacity at each power plant to serve as a baseline against which to compare the performance of the other models. That is, I separately estimate for each power plant the time-series model:

$$Size_t = \alpha + \xi_t \tag{1.7}$$

where α is the plant mean and ξ_t is the prediction error.

The next two model specifications use a subset selection procedure. For each specification, I estimate a model for each pair of power plant and weather station and then, for each power plant, select the model that minimizes the cross-validated mean squared error (MSE). For each pair of power plant and weather station, the model is:

$$Size_t = \alpha_{hm} + f(Temp_{st}) + \mathbf{X}_t\boldsymbol{\rho} + \xi_t \tag{1.8}$$

where α_{hm} represents a set of indicators for each hour of day (h) within a month of the year (m), $Temp_{st}$ is the time series of temperatures at weather station s, $f()$ is estimated nonparametrically using 2.5°C (4.5°F) temperature bins, $\mathbf{X}_t$ is an additional set of potential explanatory variables, and ξ_t is again the prediction error. In the first specification, I consider no variables beyond temperature. In the second specification, I include system-wide demand and lagged local market capacity in $\mathbf{X}_t$, up to a third-order term for each variable. I use LASSO to select which of these additional explanatory terms to keep in the model for each pair of power plant and weather station. For each power plant, I select the two models — one for each specification — that minimize the cross-validated MSE. This matches each power plant to the single weather station that is most predictive of the plant's local market capacity.

The final two model specifications include temperatures from all 75 weather stations in a single model for each power plant, and I use LASSO to select which temperatures to keep in the model. That is, I separately estimate for each power plant:

$$Size_t = \alpha_{hm} + \mathbf{Z}_t\boldsymbol{\zeta} + \xi_t \tag{1.9}$$

where α_{hm} is as above, Z_t is the set of all explanatory variables, and ξ_t is again the prediction error. In the first of these specifications, Z_t includes only temperatures, up to a third-order term for each of the 75 weather stations and an interaction for each pair of weather stations. In the second, I again add system-wide demand and lagged local market capacity; I include up to a third-order term for each variable, the interaction of the two, and the interactions of each one with each temperature measurement. Each specification has at least 3,000 possible explanatory variables, in addition to the indicators, and I use LASSO to select which terms to keep in the model for each power plant.[42] This procedure yields two additional prediction models — one of each specification — for each power plant.

[41] I stratify each fold by hour of day within a month of the year to ensure each fold sufficiently captures the seasonality and daily cycles present in electricity market data.

[42] When using LASSO to select terms to include in the model, I use the one standard error rule to select the value of the tuning parameter, typically denoted as λ, that scales the shrinkage penalty. That is, I consider

Table 1.2 summarizes the results of these five models. I calculate each metric at the plant level, and the table displays the mean and standard deviation of all power plants. Column (1) corresponds to the baseline model of Equation (1.7), Columns (2) and (3) correspond to the subset selection models of Equation (1.8), and Columns (4) and (5) correspond to the LASSO models of Equation (1.9). The subset selection models perform better across all measures than the comparable LASSO models — that is, Column (2) performs better than Column (4), and (3) better than (5). This result suggests that the relationship between temperature and local market capacity is sufficiently nonparametric that a flexible function of a single weather station's temperature is a better predictor of local market capacity than is the full set of temperatures with higher-order terms and interactions. For this reason, I focus on the subset selection models that match each power plant to a single weather station.

Figure 1.7 plots the distribution of prediction errors from the baseline and both subset selection models. The lightest distribution shows the variation that exists in local market capacity, which is bimodal due to the mass of large negative values, corresponding to hours when a plant faces a local market that is much smaller than its average. The middle distribution, which plots the errors from the model in Column (2) of Table 1.2, shows that a single, nonparametric temperature and temporal indicators account for roughly one-third of the variation in local market size, with much more mass at the center of the distribution. The further addition of more explanatory variables increases this predictive power to nearly 45% of baseline variation, as shown in the darkest distribution and Column (3) of Table 1.2. While these simple models do not perfectly predict the size of the local market in which a power plant will operate, they account for a sufficient amount of variation in local market size to be useful in both the first stage of my 2SLS estimation and providing a simple heuristic that even the less sophisticated firms could use when operating in this market.

The main result that I take from this prediction model to my 2SLS regression specification is the matching of each power plant to the weather station that best predicts its local market capacity. Although both system-wide demand and lagged local market capacity add predictive power to the time series model for each power plant, I do not include them in my panel 2SLS regression for econometric reasons.[43] Thus, I use the matched weather stations from the subset selection model, summarized in Column (2) in Table 1.2, that includes temperature but no additional explanatory variables.

I instrument for a power plant's local market size with the temperature at its matched weather station while controlling for the local temperature near the power plant and, hence, the temperature-dependent characteristics of the power plant's local market. This approach, however, requires that these temperatures are only weakly correlated so the matched temperature has sufficient variation that is orthogonal to that of the local temperature. Figure

many possible values of the tuning parameter and, for each value, calculate the mean and the standard error of the cross-validated MSE, where the mean and standard error are taken over the five folds of data. I consider all tuning parameter values that yield a mean MSE within one standard error of the minimum MSE. From this set of tuning parameters, I select the value that yields the models with the fewest number of variables. This common procedure produces the simplest model with a mean MSE that is statistically indistinguishable from the minimum MSE.

[43] I include hour of sample fixed effects, which are collinear with system-wide demand. I also include unit fixed effects, so I omit lagged local market capacity to avoid dynamic panel concerns.

1.8 plots two histograms of the distance from power plants to weather stations, one for the closest weather station to each power plant, and the other for the matched weather station. The majority of power plants have a weather station within 25 km (15.5 miles), so the temperature at the closest weather station is a good representation of the temperature at the plant and in the plant's local market. A matched weather station, however, can be more than 800 km (500 miles) away from the power plant, and the average distance from a power plant to its matched weather station is roughly 305 km (195 miles). These large distances suggest that, even when controlling for local temperature, the matched temperature has sufficient variation that is correlated with the plant's local market size but orthogonal to other local market conditions.

Figure 1.9 plots the distribution of matched temperatures, as both a kernel density and a histogram by temperature bins. These distributions depict an important feature of the Texas electricity market: high temperatures are common. The median temperature is 22.7°C (72.9°F), and the modal temperature bin is just above 25°C (77°F). At the upper end of the temperature distribution, temperatures greater than 30°C (86°F) occur in 21.5% of observations, and temperatures exceed 35°C (95°F) in 7.5% of observations. These high temperatures drive transmission congestion and local market formation, as described in Section 1.4, and the prevalence of these high temperatures underscores the importance of local markets in this setting.

1.5.2 Empirical specification

I estimate the causal effect of market size on markups by estimating the following two equations with a 2SLS regression:

$$Markup_{it} = \beta \widehat{Size}_{it} + g(LocalTemp_{it}) + \gamma_{ihm} + \delta_t + \varepsilon_{it} \tag{1.10}$$

$$Size_{it} = f(MatchedTemp_{it}) + g(LocalTemp_{it}) + \eta_{ihm} + \theta_t + \omega_{it} \tag{1.11}$$

where the second stage in Equation (1.10) uses the fitted values of $Size_{it}$ — the market size facing unit i at hour t — estimated by the first stage in Equation (1.11). The parameter of interest is β, which gives the relationship between this fitted value of market size and $Markup_{it}$, the markup of price over marginal cost.

In the first stage I estimate a model of market size as a function of $MatchedTemp_{it}$, the average temperature during hour t at unit i's matched weather station. I consider two different functional forms of temperature. In the more flexible specification, I nonparametrically estimate market size as a function of temperature using 2.5°C (4.5°F) temperature bins, as in the prediction models using subset selection. In the other specification, I estimate market size using a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). When reporting 2SLS results, I indicate which first-stage specification was used.

I flexibly control for the temperature during hour t at unit i's closest weather station, $LocalTemp_{it}$, in each stage. I again use 2.5°C (4.5°F) temperature bins. This function nonparametrically controls for conditions in the local market that are influenced by temperature, such as the level of demand due to indoor cooling and heating, as well as any potential ef-

fects of temperature on power plant operations, including reduced efficiency due to extreme temperatures.[44]

I include in each stage a set of unit fixed effects, γ_{ihm} and η_{ihm}. Electricity markets exhibit cyclical behavior over the course of a day, with demand and wholesale price often peaking in the afternoon and reaching a low in early morning hours. Electricity markets also exhibit seasonality, with demand and wholesale price often at their highest during summer months, and this is particularly true of the Texas electricity market due to a large demand for air conditioning at high temperatures. To control for these regular patterns in the market and the manner in which firms may respond to them at each power plant, I allow for a different unit-specific effect for each hour of day (h) within a month of year (m). I also include in each stage a set of hour-of-sample fixed effects, δ_t and θ_t, to control for contemporaneous shocks common to all units.

After controlling for the local temperature and these fixed effects, the identifying variation is within-unit variation in deviations from market-wide averages. For example, isolated weather patterns that induce not only deviations in the mean temperature across the market but also deviations in the distribution of temperatures across the market. Or, in the case of market size, transmission congestion that results in not only heterogeneous local market sizes but also deviations from the typical grid topology.

The remaining terms, ε_{it} and ω_{it}, are idiosyncratic errors. These errors are likely to be highly correlated over time. For example, an unobserved shock that affects a unit's market size or markups in one hour often persists into the future. These errors are also likely to be highly correlated over space because all units within a local market face the same local market size. To account for this correlated error structure, I two-way cluster standard errors at the unit and date levels, which allows for arbitrary serial correlation within a unit and arbitrary cross-sectional correlations within the same date. This level of clustering provides the correct inference so long as one unit's error term is not correlated with the error term of a different unit on a different date.

In order to interpret my estimate of β in Equation (1.10) as the causal effect of market size on markups, the standard assumptions of 2SLS must hold:

$$\text{Cov}(MatchedTemp_{it}, Size_{it} \mid LocalTemp_{it}, \eta_{ihm}, \theta_t) \neq 0 \tag{1.12}$$

$$\text{Cov}(MatchedTemp_{it}, \varepsilon_{it} \mid LocalTemp_{it}, \gamma_{ihm}, \delta_t) = 0 \tag{1.13}$$

The relevance assumption in Equation (1.12) ensures that variation in the matched temperature captures some of the variation in market size; I have already shown this to be true in the prediction models and will address it again when discussing results. The exclusion restriction in Equation (1.13) ensures that, conditional on the local temperature and high-dimensional fixed effects, the effect of matched temperature on markups acts only through market size and through no other channels.

As with any instrumental variables approach, the primary threat to identification is the validity of the exclusion restriction. In this case, the exclusion restriction would be invalid if a power plant's matched temperature is correlated with the plant's markup in a way that

[44] Extreme heat at the closest weather station has a small effect on plant efficiency, whereas extreme heat at a plant's matched weather station has no effect on efficiency. I show these results in Appendix A.4.

is not explained by the local market size, the local temperature, the plant's typical daily or seasonal patterns, or a shock common to all plants in the full market. For example, the consumers in the plant's local market would have to adjust their demand for electricity in response not to the local temperature, but rather to the temperature at the geographically distant matched weather station, which is unlikely. These controls leave no clear channel, other than local market size, through which the matched temperature could be correlated with the markup decision at the power plant.

1.6 Results

1.6.1 First-stage results

I consider three different measures of market size: the production capacity of the local market, the number of units in the local market, and the number of firms in the local market. I first show how temperature affects market size by plotting the first-stage results of Equation (1.11) for each of these size measures. The results for capacity, units, and firms are depicted in Figure 1.10. The points and error bars show estimates and 95% confidence intervals for the binned specification;[45] the curve and shaded area show the estimate and 95% confidence interval for the natural cubic spline. In each specification, I estimate effects relative to a temperature of 16.25°C (61.25°F),[46] which roughly corresponds to the midpoint of the temperatures I observe. Because results are relative to this reference temperature, negative values on the figures do not imply local markets with negative size, but rather local markets that are smaller than a comparable market at the reference temperature.

Local market capacity, the number of units in the local market, and the number of firms in the local market are highly correlated, so their temperature-response functions share a similar shape in Figure 1.10, although the units and scale of the y-axis vary. For all three measures, the effect on market size is relatively flat and precisely estimated throughout the middle range of temperatures, and even at the lowest temperatures in most cases. Much of the temperature mass lies within this range, as plotted in Figure 1.9, so a unit's matched temperature has no marginal effect on its market size for many hours of my sample. When temperatures exceed 30–35°C (86–95°F), however, the average local market size faced by a power plant is significantly reduced. At the extreme, when temperatures exceed 40°C (104°F), local markets decrease in size by 6.5 GW, 15 units, or 4 firms, as compared to a moderate temperature. This represents a reduction of 9–12% from the mean local market size.

[45] All interior bins are 2.5°C (4.5°F) wide. The first and last bins correspond to temperatures below -10°C (14°F) and above 40°C (104°F), respectively. Point estimates are placed at the midpoint of each bin and give the average response to a temperature within that bin.

[46] In this analysis, the fixed effects allow for each unit and each hour of sample to have a different intercept, and I estimate the average temperature-response function relative to those intercepts. One temperature must be selected as the reference temperature, or the point at which the intercept is calculated. This choice is arbitrary and innocuous. If I selected a different temperature to be the reference, the temperature-response function would simply shift up or down to intersect the x-axis at that temperature.

These first-stage results confirm the general intuition described in Section 1.2 and the aggregate data presented in Section 1.4. It is worth considering, however, exactly what is being identified in these regressions. Because of the empirical specification and the identifying variation that remains after controlling for local market conditions with local temperature and for unobservables with fixed effects, these results do not imply that local markets only form at extreme temperatures or that local market sizes vary by only 6.5 GW, 15 units, or 4 firms. Local markets do form at more moderate temperatures, as shown in Figure 1.6, but the size of these local markets are fully explained by a combination of the local temperature near the power plant and the fixed effects, so these observations provide no identifying variation in the first stage.

Instead, these first-stage regressions are identified off of variation that differs from the local temperature and hourly or unit-level averages. For example, when one unit's matched weather station experiences a temperature greater than 40°C (104°F), most weather stations throughout the state also experience high temperatures, so transmission congestion is likely to be pervasive during that hour. All units face relatively small local markets, but the first-stage regression identifies that units matched to the most extreme temperatures face the smallest local markets. This is precisely the innovation of my research strategy. I exploit panel variation in market size, not simply time series variation, so I can causally estimate how firms in the smallest local markets markup their electricity relative to firms in larger local markets during the same hour.

1.6.2 Reduced-form results

I look at the markups submitted at three quantities of production: 85%, 75%, and 65% of nameplate capacity.[47] I show how matched temperature affects markups at each of these quantities through the reduced-form regression:

$$Markup_{it} = f(MatchedTemp_{it}) + g(LocalTemp_{it}) + \upsilon_{ihm} + \phi_t + \psi_{it} \tag{1.14}$$

where the terms in Equation (1.14) are equivalent or comparable to those in Equations (1.10) and (1.11).

The results of these reduced-form regressions are plotted in Figure 1.11.[48] The effect on markups is relatively flat and not statistically different from zero at temperatures below 30°C (86°F), and this is true for all three quantities. When temperatures exceed 30–35°C (86–95°F), however, markups increase significantly. At the extreme, when temperatures exceed 40°C (104°F), markups at 85% capacity are as much as $220 per MWh higher than during comparable hours at a moderate temperature; this effect of extreme heat is substantial — more than six times as large as the average marginal cost for a fossil fuel-fired power plant and nearly twice as large as the average markup at this quantity. Similarly, markups at 75%

[47] See Sections 1.2 and 1.4 for a description of offer curves and an explanation for focusing on these three points on the curve.

[48] As in Figure (1.10), the points and error bars show estimates and 95% confidence intervals for the binned specification, and the curve and shaded area show the estimate and 95% confidence interval for the natural cubic spline. I again estimate effects relative to a temperature of 16.25°C (61.25°F), so effects are given not in absolute levels but relative to a comparable hour at the reference temperature.

and 65% of nameplate capacity are as much as roughly $140 and $120 per MWh higher, respectively, when temperatures exceed 40°C (104°F), relative to moderate temperatures; the effect at each of these quantities is 1.35 and 1.65 times as large the average markup at the respective quantity. The range of temperatures that induces these large effects on markups corresponds to temperatures that cause local markets to shrink, further suggesting a strong correlation between smaller market size and higher markups.

1.6.3 Results for local market capacity

I estimate the causal effect of market size on markups using 2SLS as described in Equations (1.10) and (1.11). I first consider market size measured by local market capacity, and the results are shown in Table 1.3. This table reports results for markups at all three quantities; for each quantity level, the table reports three regressions: OLS, 2SLS using the temperature bin first-stage specification, and 2SLS using the natural cubic spline first-stage specification. Looking first at markups at 85% capacity, the OLS regression finds no effect of local market capacity on markups; the estimate is small and not statistically significant. The 2SLS estimates, however, are large and indicate markups are higher when markets are smaller. For every one less gigawatt of power plant capacity in the local market, a firm marks up prices by as much as an additional $25 per MWh. This result implies that a reduction in local market capacity equal to 10% of mean local market capacity causes markups to increase by up to $135 per MWh. This effect is large — more than 3.5 times the average marginal cost of a fossil fuel-fired power plant and nearly 1.2 times the average markup I observe at 85% capacity.

A more than doubling of markups for a relatively small change in local market capacity may seem extreme, but it is again useful to consider the identifying variation for this estimation. These 2SLS regressions are identified off of variation in local market capacity caused primarily by hours of extreme heat. During these hours, the grid is likely stressed due to congested transmission lines and high levels of demand. When this situation occurs, most local markets are operating on the inelastic portions of their aggregate offer curves, so firms face inelastic residual demand and have strong incentives to exercise market power by marking up prices. If a local market in this scenario becomes even smaller, the residual demand facing every unit becomes substantially more inelastic, strengthening the incentives for every firm to exercise market power and yielding large increases in markups.

While these hours of extreme heat events represent only a minority of the hours when this market operates, they correspond to some of the most challenging hours for the operation of the electrical grid, and extreme levels of market power may exacerbate these challenges. Additionally, these times of extreme heat will become more common in the face of climate change, which I explore in Section 1.7. By employing an empirical strategy that exploits variation during these hours, this analysis contributes to better understanding one of the driving factors of market power during these critical times.

Returning to the results in Table 1.3, the effects of local market size on markups at 75% and 65% of capacity are similar to those at 85% of capacity, although effect sizes are smaller in absolute value at each subsequently smaller quantity. These results show that, as local market capacity shrinks, not only do markups increase but also offer curves become

more steep.[49] Thus, when a local market is smaller, firms face fewer competitors, and those competitors submit offers that are less elastic. Both of these factors contribute to a firm facing less elastic residual demand, which increases the firm's incentive to further markup prices. These regressions estimate the combined effect of these two factors.

Across all three quantities, the OLS and 2SLS estimates are vastly different. The OLS results show that markups are uncorrelated with the size of the market ultimately faced by the firm, which is consistent with local market size being, on average, unpredictable by firms. As shown in Section 1.5, however, a portion of the variation in local market size is predicted by temperature. By instrumenting for market size with matched temperature in the 2SLS regressions, I isolate the variation in market size that a firm is able to predict when submitting offer curves for its power plants. As a result, the 2SLS regressions estimate a local average treatment effect (LATE). That is, they estimate the effect on markups of a predictable change in local market capacity that is induced by temperature, but not the effect of a change in local market capacity due to a different and potentially unpredictable cause.

One concerning set of results in Table 1.3 is the first-stage F-statistic for each 2SLS estimation. The largest F-statistic results from using the natural cubic spline first-stage specification for markups at 85% capacity, shown in Column (3). But this F-statistic is only 6.47, which falls short of the F-statistic "rule of thumb" of 10 (Staiger and Stock 1997), indicating the relevance assumption of Equation (1.12) may not be sufficiently strong and these 2SLS estimates may be biased due to a weak instruments problem. Using the Stock and Yogo (2005) test for 2SLS bias due to weak instruments, I reject that the bias of 2SLS relative to the bias of OLS is greater than 30% for the spline specifications, which is less than one standard error of my estimate. Additionally, the direction of the bias is toward the OLS estimate, so 2SLS estimates are biased toward zero. Thus, I interpret these 2SLS estimates as conservative estimates of the true LATE with only a small potential bias. For the remainder of this analysis, however, I focus on results for markups at 85% capacity using the natural cubic spline first-stage specification, which uniformly exhibit the largest first-stage F-statistic.[50]

1.6.4 Results for number of units or firms in local market

Tables 1.4 and 1.5 display the results of regressions using the number of units in the local market and the number of firms in the local market, respectively, as the measure of market size. As with the first-stage results in Figure 1.10, these regression results are qualitatively consistent with those for local market capacity because all three measures of market size are highly correlated, although the estimates differ to account for the scale of the market size measure being considered.

[49] I assume a unit's marginal cost is constant in quantity, so larger markup effects at larger quantities must be due to larger increases in the offer price — that is, an offer curve that becomes more steep in response to less local market capacity. I directly estimate effects on offer curves in Appendix A.4 and show this to be true.

[50] I present results for markups at 75% and 65% capacity using the natural cubic spline first-stage specification and for markups at 85% capacity using the binned first-stage specification in Appendix A.4.

Focusing on effects at 85% capacity, a firm increases markups by an additional \$10 per MWh for every one fewer unit competing in a local market and by more than \$40 per MWh for every one fewer firm competing in the local market, as shown in Column (3) of each table. These estimates imply that a firm increases markups by \$155–180 per MWh if the size of the market shrinks by 10% of the mean market size. These increases in markups are even larger than for local market capacity — more than four times the average marginal cost of a fossil fuel-fired power plant and 140–160% of the average markup I observe at 85% capacity. The effects on markups at smaller quantities are similarly comparable to those for local market capacity.

1.6.5 Decomposition of effects

The results in Tables 1.3–1.5 give the average effect of market size on markups. As shown in the comparative statics of Section 1.3, however, different types of capacity and units may have vastly different effects on markups. For example, Equation (1.5) predicts potentially opposite effects on a plant's markups depending on the ownership of local market capacity — more competing capacity reduces markups, whereas more capacity owned by the plant's own firm may increase the incentive to mark up offer prices. Similarly, Equation (1.6) predicts different magnitudes of effects depending on whether capacity in the local market is dispatchable or not, with dispatchable capacity yielding larger effects in absolute value.[51]

To test these predictions, I regress markups on two market size variables. In the first case, these variables are local market capacity owned by competing firms and local market capacity owned by the plant's own firm; in the second, dispatchable capacity and nondispatchable capacity. I again use a 2SLS approach and estimate each size variable as a function of matched temperature in the first stage. Table 1.6 displays the results of regressions to test these predictions on markups at 85% capacity.[52] Column (1) duplicates the main regression result from Column (3) of Table 1.3 to serve as a reference.

Column (2) of Table 1.6 shows the effect on markups by ownership, which confirms the predictions of the model. Less competing capacity in a local market causes markups to increases at nearly double the rate of the average effect, although these point estimates are not statistically different from one another. Conversely, and as predicted, markups increase when there is more capacity in a plant's local market that is owned by the plant's firm, and the size of this effect is an order of magnitude larger than the effect of competing capacity.

Column (3) of Table 1.6 shows the effect on markups by unit type. Dispatchable capacity, which contributes more to the slope of the residual demand curve, has an effect on markups that is roughly equivalent to the average effect. Nondispatchable capacity, however, has no statistically distinguishable effect on markups. These results confirm the results of the

[51] I define nondispatchable capacity as nuclear and renewable plants, and the remaining capacity as dispatchable. The nondispatchable plants almost always submit offer curves at negative prices and with approximately no slope at the quantities where the plants typically operate, whereas the dispatchable plants typically submit offer curves with prices that increase in quantity. See Figure 1.4 for example offer curves and Figure 1.5 for the mean offer curve by plant types.

[52] Results for markups at other quantities, as well results using number of units or firms in the local market, are in Appendix A.4.

model, that capacity with less elastic offer curves have a larger effect on markups because the offer curves have a larger effect on the elasticity of residual demand.

1.6.6 Robustness checks for local controls and temperature matching

An important feature of my identification strategy is matching each power plant to a weather station that is geographically far from the power plant, so I can also control for the temperature at the weather station closest to the power plant. This approach, however, has two potential concerns. First, local temperature may not be sufficient to fully control for conditions in the plant's local market, particularly when that local market is large. Second, if there is insufficient variation in matched temperature after controlling for local temperature and fixed effects, then estimates may be attenuated. To explore these concerns, I use alternate local controls and matched weather stations.

I first consider different local controls. The primary reason to include local temperature is to control for the level of demand in the local market. Higher levels of demand cause the local market to clear further out on the aggregate offer curve, which corresponds to less elastic supply and may affect markups independent of local market size. To directly control for where the local market clears, I include a higher-order function of local capacity factor — the ratio of production to capacity in the local market — as a control, both instead of and in addition to local temperature. To address the second concern, I also consider a model with no local controls. Table 1.7 displays the results of these regressions for the effect of local market capacity on markups at 85%.[53] The estimated effect changes little and no estimates are statistically different from one another.

Another method to address both concerns is to require a power plant's matched weather station to be a given distance away from the plant.[54] A minimum matching distance both increases the likelihood that a plant's matched temperature is outside its local market and decreases the correlation between matched and local temperature, to ensure more temperature variation remains when controlling for local temperature. Table 1.8 displays the results when requiring a plant's matched weather station to be 100, 200, or 300 km (62, 124, or 186 miles) away from the plant. The effect of local market capacity on markups at 85% capacity is not statistically different at any of these minimum matching distances.

1.6.7 Matched temperature or local temperature

An important innovation of this analysis is the use of a prediction model to match each power plant to the weather station that is most predictive of the plant's local market capacity, rather than simply using the local temperature closest to the power plant. Not only is this matching procedure necessary for causal identification — instrumenting with local temperature does not satisfy the exclusion restriction, as discussed previously — but it also

[53] See Appendix A.4 for results using the other measures of local market size and other quantities.

[54] I enforce no such requirement in the original match, even allowing a power plant to be matched to its closest weather station. Only three plants are originally matched to the closest weather station.

exploits the variation that firms are more likely to consider when submitting offer curves. As a result, 2SLS regressions using matched temperature should yield less bias toward zero, as compared to estimates that use local temperature. To see the importance of this strategy and its effects on my estimates, I also perform 2SLS regressions using local temperature, rather than matched temperature, as the instrument for local market size.

Table 1.9 shows the results of these regressions; Columns (1)–(3) duplicate previous results using matched temperature, and Columns (4)–(6) show the estimates from comparable regressions but with local temperature as the instrument. For all three measures of market size, the estimates when instrumenting with local temperature are roughly one-fourth of the estimates when instrumenting with matched temperature, and the estimates using local temperature are not statistically significant. These results show that my empirical strategy is necessary not only to find the causal effect of market size on markups but to find any empirical relationship between market size and the exercise of market power, as predicted by theory, that is not subject to large attenuation bias.

1.7 Welfare effects

The exercise of market power through increased markups has two main effects on welfare in this setting: large transfers from consumers to producers and a deadweight loss of consumer surplus. Increased markups yield higher prices paid by the purchasers of wholesale electricity, primarily retail providers that purchase electricity to distribute to their end-use consumers. The retail electricity prices paid by the consumers are contracted in advance and rarely vary with the wholesale price in the real-time market. Thus, in the short run, increased markups cause transfers from electricity retail providers to producers of electricity but no change in the quantity of electricity produced and consumed.

In the long-run, however, retailers pass their increased wholesale costs on to consumers in the form of higher retail prices. Thus, in equilibrium, average retail electricity prices are higher in this setting than they would be in a competitive market, even though wholesale prices are not passed on to end-use consumers in real time. Higher average retail prices have two effects. First, they yield transfers from consumers to retailers and then, ultimately, to the producers of electricity. Second, they produce a long-run demand response with less electricity consumed than in a competitive counterfactual. The transfers do not create an inefficiency in this market but may be of interest for equity concerns. The reduction in consumption, however, creates a deadweight loss of consumer surplus.

I calculate the size of the transfers and deadweight loss due to increases in markups that are induced by transmission congestion and smaller local markets. I then calculate the extent to which these welfare effects will increase as a result of climate change and the higher prevalence of transmission congestion.

1.7.1 Consumer surplus and transfers to producers

To calculate the welfare effects of this source of market power, I consider the counterfactual in which firms submit offer curves as if temperature-induced transmission congestion did not

reduce the size of local markets. The first-stage results in Figure 1.10 show the extent to which a plant's local market size varies as a function of temperature at the plant's matched weather station. I focus on temperatures greater than 30°C (86°F), which correspond to the matched temperatures at which firms change their markups, as shown in the reduced-form results in Figure 1.11.

I use the first-stage estimates from the natural cubic spline specification to calculate how much larger each plant's local market capacity would be in the absence of temperature-induced transmission congestion — that is, the local market capacity at a moderate temperature,[55] as compared to the observed matched temperature. At these moderate temperatures, local markets are, on average, 1.1 GW larger. There is great heterogeneity, however, and when compared to the most extreme matched temperatures, local market capacity can be as much as 7.7 GW larger.

If firms did not consider local market size when submitting offer curves, markups would be smaller during these hot hours. Using the 2SLS estimates from Column (3) in Table 1.3, I calculate that markups would be $27.1 per MWh smaller on average, and some markups would be as much as $195.5 per MWh smaller. These smaller markups would yield lower LMPs at every power plant; I assume the reduction in LMPs at each power plant is equal to the reduction in its markup. Due to these lower markups, and their resulting lower prices, the total value of the electricity transacted in the real-time wholesale electricity market during my study period is $8.3 billion less in this counterfactual. This value represents only the lower prices paid by electricity retailers in the wholesale market, which they pass on to consumers, and not a change in the quantity of electricity produced or consumed. Thus, the total transfer from consumers to producers due to this exercise of market power is $2.1 billion annually.

These lower wholesale prices would yield lower average retail prices paid by end-use consumers. Over the four years of my study period, nearly 1.3 billion MWh of electricity was consumed. Thus, without this source of market power, the average retail price of electricity would be $6.5 per MWh lower. During these years, consumers in the state of Texas paid an average of $87.9 per MWh for electricity,[56] so roughly 7.4% of the average retail price is due to this source of market power. Recent estimates of the long-run elasticity of demand for electricity range from -0.09 to -0.27 (Deryugina, MacKay, and Reif 2017).[57] Assuming a constant elasticity of demand for electricity, these elasticity estimates imply electricity consumption is 0.69–2.09% higher in the counterfactual without these markups. Thus, as compared to this counterfactual, consumers in this market lose $7.1–21.5 million of surplus annually due to higher retail prices and a lower level of electricity consumption.

These calculations show that, in this setting, there is little welfare loss due to the exercise of market power that is induced by transmission congestion and smaller lower markets. These markups result in only $7.1–21.5 million of deadweight loss annually, in a market

[55] I use the local market size at the reference temperature of 16.25°C (61.25°F). A different moderate temperature would yield similar results because the local market capacity is not statistically different over the range 5–30°C (41–85°F).

[56] Annual average retail electricity price is from the EIA's Electric Power Annual.

[57] These elasticity estimates are for residential consumers. Commercial and industrial consumers are likely to exhibit larger elasticities, so the resulting deadweight loss is a conservative estimate.

that transacts nearly $12 billion of electricity each year. This source of market power does, however, result in large transfers from consumers to producers, $2.1 billion annually. These transfers are substantial and may be a cause for concern on equity grounds.

1.7.2 Climate change effects

Because transmission congestion is driven by ambient temperature, climate change is projected to have an important effect on the size of local markets and the resulting exercise of market power. To see this effect, I consider a counterfactual in which temperatures are 2.5°C (4.5°F) warmer than I observe during my study period, which corresponds to roughly 50 years in the future under a business-as-usual warming scenario.[58]

I first examine how climate change will alter the prevalence of transmission congestion and the number of local markets that form. I calculate the distribution of the number of local markets as a nonparametric function of mean temperature in the state. I then assign to each hour the distribution of local markets that corresponds to a temperature that is one bin warmer. Figure 1.12 shows the results of this calculation. The top panel displays histograms of the number of local markets that form in the observed data and in the counterfactual under 2.5°C (4.5°F) warming; the bottom panel shows the difference between the warming scenario and the observed data. Due to a warmer climate, the portion of time that the market is fully integrated will decrease by 2.7 percentage points. Instead, it is more likely the market will fragment into many markets. In particular, the amount of time that the market splits into 7 or more local markets will increase by 3.3 percentage points.

This increased frequency of transmission congestion results in smaller local markets and, hence, increased market power through markups. Again using the first-stage estimates from the natural cubic spline specification, as shown in Figure 1.10, I calculate that power plants will face local markets that are 0.3 GW smaller on average, due to warming. These effects are not linear, however, and for power plants with matched weather stations that already experienced extreme temperatures, climate change will reduce the capacity of their local markets by as much as 1 GW in some hours. Using the 2SLS estimates from Column (3) in Table 1.3, I calculate that markups will increase by $7.1 per MWh on average, and some markups will be as much as $26.0 per MWh larger.

These increased markups will yield higher LMPs; I again assume the increase in each power plant's LMP is equal to the increase in the plant's markup. Due to these higher prices, an additional $2.5 billion would be transacted annually in the real-time wholesale electricity market if the same amount of electricity was consumed. In the long run, however, these increased markups will result in average retail prices that are $7.8 per MWh, or 8.8%, higher than prices during 2011–2014. These increased retail prices will yield less consumption; using the same range of elasticities as above, consumption will decrease by 0.76–2.26%.[59]

[58] 2.5°C (4.5°F) warming by 2065, as compared to 2011–2014, is roughly consistent with temperature projections under RCP 8.5 (Riahi et al. 2011). Because I use 2.5°C (4.5°F) temperature bins in my nonparametric temperature-response functions, this warming is equivalent to counterfactual temperatures that are one temperature bin warmer than what I observe in the data.

[59] I consider the effect of climate change only on the supply and transmission of electricity, not on the demand for electricity. See Auffhammer, Baylis, and Hausman (2017) for estimates of how climate change is

This reduction in consumption corresponds to a consumer surplus loss of $9.4–27.8 million annually.

Thus, climate change will yield a deadweight loss of $9.7–28.8 million annually because of the effect of warmer temperatures on local market formation and the resulting exercise of market power by firms. This value is roughly equivalent to the deadweight loss due to the same source of market power during the study period. Climate change will also generate large transfers from consumers to producers, $2.4–2.5 billion annually, as a result of increased markups, which is again comparable in size to what I calculate for the years 2011–2014. Thus, both deadweight loss and transfers from consumers to producers, which result from higher markups due to temperature-induced changes in local market size, will roughly double as a result of climate change. An important caveat is that this analysis considers the effect of warmer temperatures on the supply and transmission infrastructure present in 2011–2014, and much of this infrastructure will be updated or replaced within the next 50 years. These results highlight the importance of considering climate change when making these long-run planning and investment decisions.

1.8 Conclusion

This paper studies the effect of market size on market power in the Texas electricity market. I exploit a novel source of exogenous variation in market size, transmission congestion due to ambient temperature shocks, to estimate how the size of the market in which a power plant operates affects its markups. I find that a 10% reduction in market size causes markups to more than double, compared to mean values. These results show that even a relatively small change in market size can greatly exacerbate the ability of a firm to exercise market power in this setting. These markups yield a small loss of consumer surplus, $7.1–21.5 million annually, because demand for electricity, even in the long run, is relatively inelastic. These markups do, however, generate large transfers from consumers to producers, $2.1 billion annually, which may be a cause for concern on equity grounds.

This work makes an important contribution to the literature by estimating the effect of market size on market power using weaker assumptions than prevailing approaches. Transmission congestion due to ambient temperature shocks provides an exogenous source of variation that I use to estimate this effect without making structural assumptions on firm behavior or the nature of competition. It is, however, an hourly source of variation, meaning my estimates are short-run effects. In this setting, short-run effects are particularly important because supply must perfectly meet demand in real time. While these estimates may not be directly transferable to other contexts, the direction and magnitude of the effects give important information about how sophisticated firms compete in a dynamic market with large stakes.

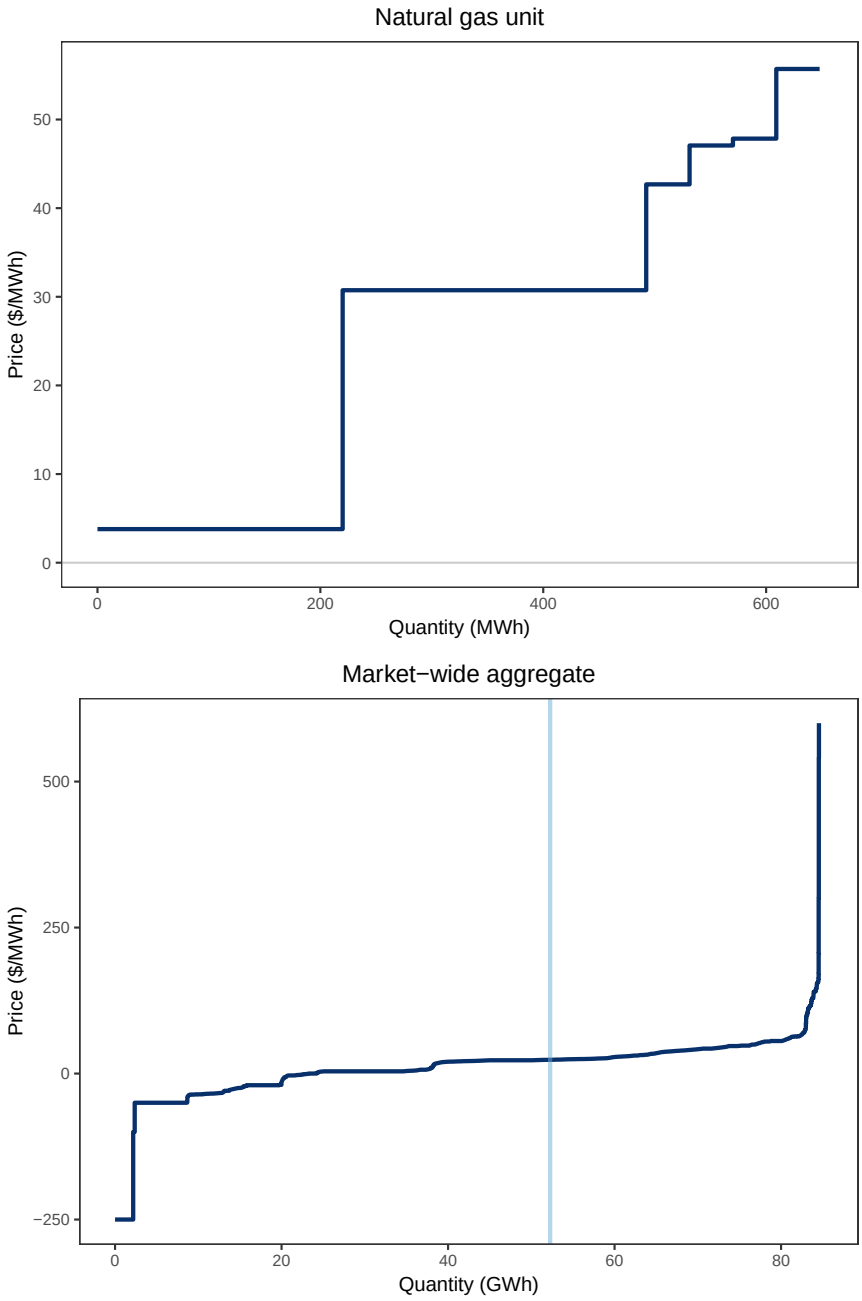
This distinction between the short-run and long-run effects of market size on market power points to an important area of future research. In electricity markets, increased

expected to increase electricity consumption. At higher levels of electricity demand, the effects of increased markups are larger, so my calculations provide conservative estimates of the welfare effects of climate change in this setting.

markups due to transmission congestion may lead to new investments in the long run. Firms may site new power plants in areas of the grid that often experience small local markets, which would decrease the market power of the incumbents, but this investment may not be socially optimal since market power causes relatively small deadweight loss in this setting. Understanding how this source of short-run market power affects long-run market outcomes, including trade-offs between short-run and long-run inefficiencies, is an important area of future research.

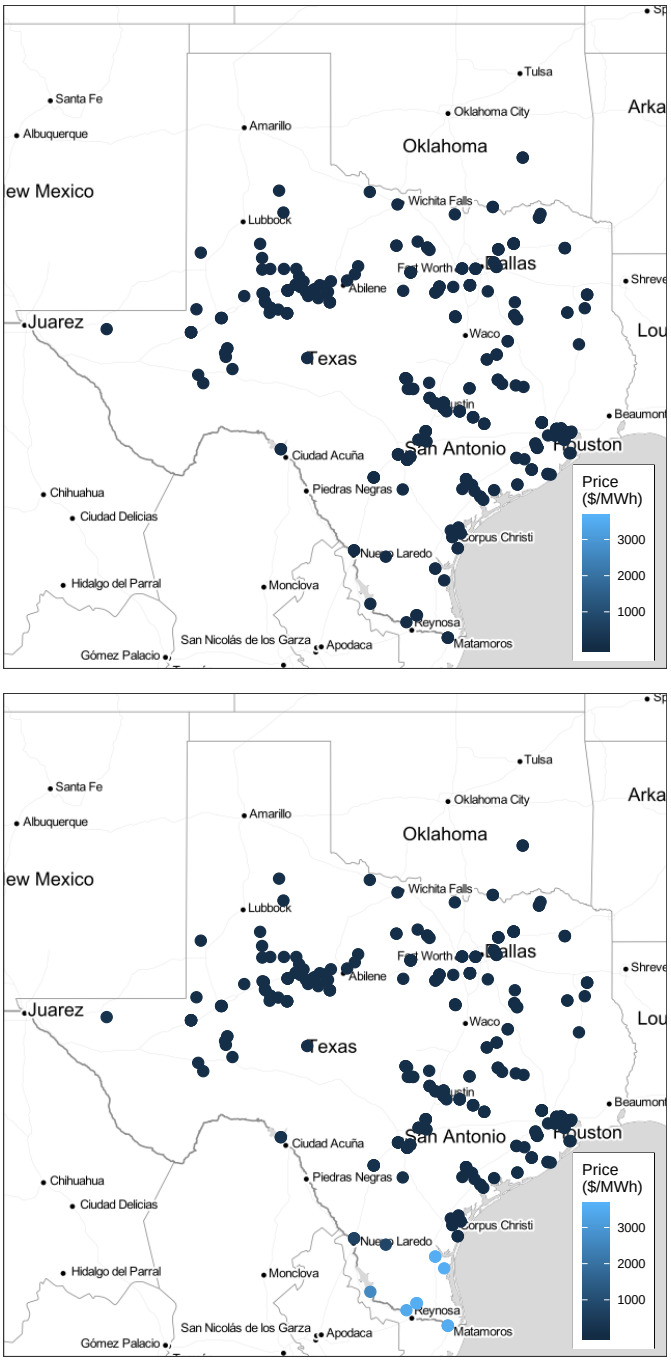
Figures

Figure 1.1: Offer curves



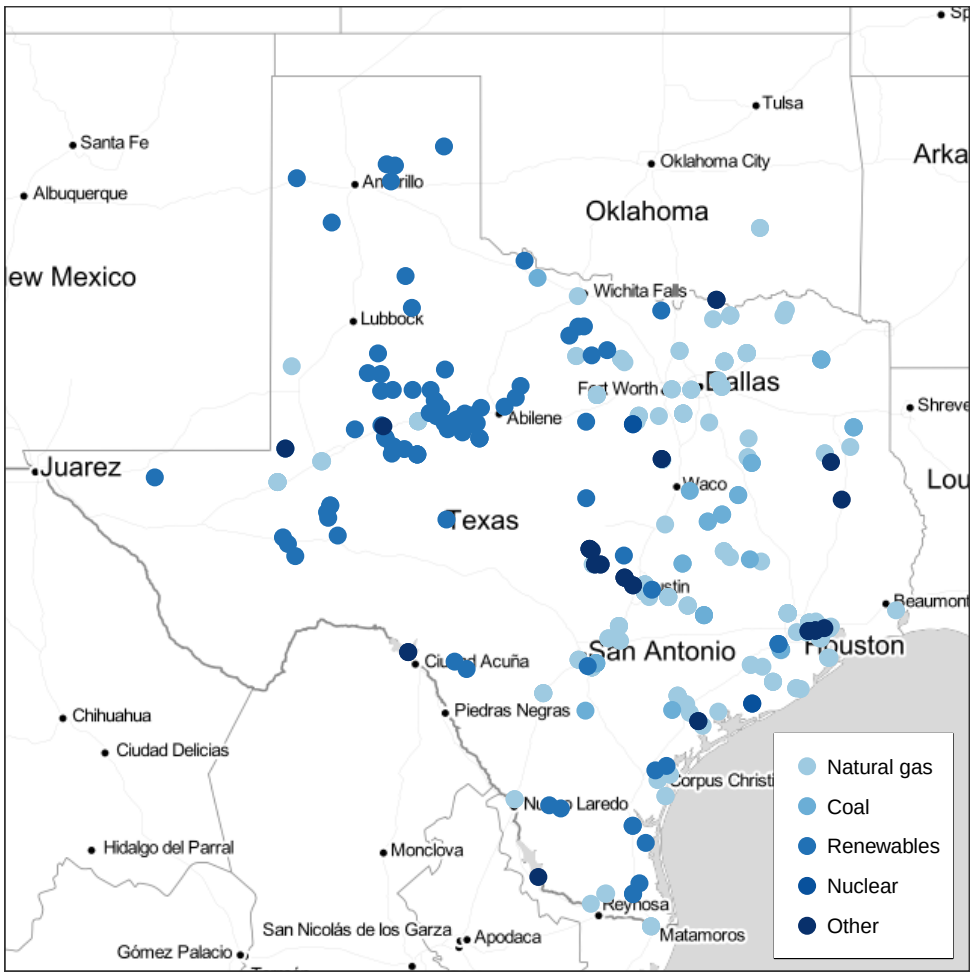
Notes: The top panel plots an example offer curve for a natural gas-fired unit. This curve specifies the amount the unit will produce at each price. The bottom panel plots a market-wide aggregate offer curve, which sums the quantity of all unit-level offer curves for each price. Note that the axes differ from the top panel. The vertical line plots the system-wide demand. In the absence of binding system constraints, the market will clear as a uniform-price auction at the intersection of the two curves.

Figure 1.2: Local markets and locational marginal prices



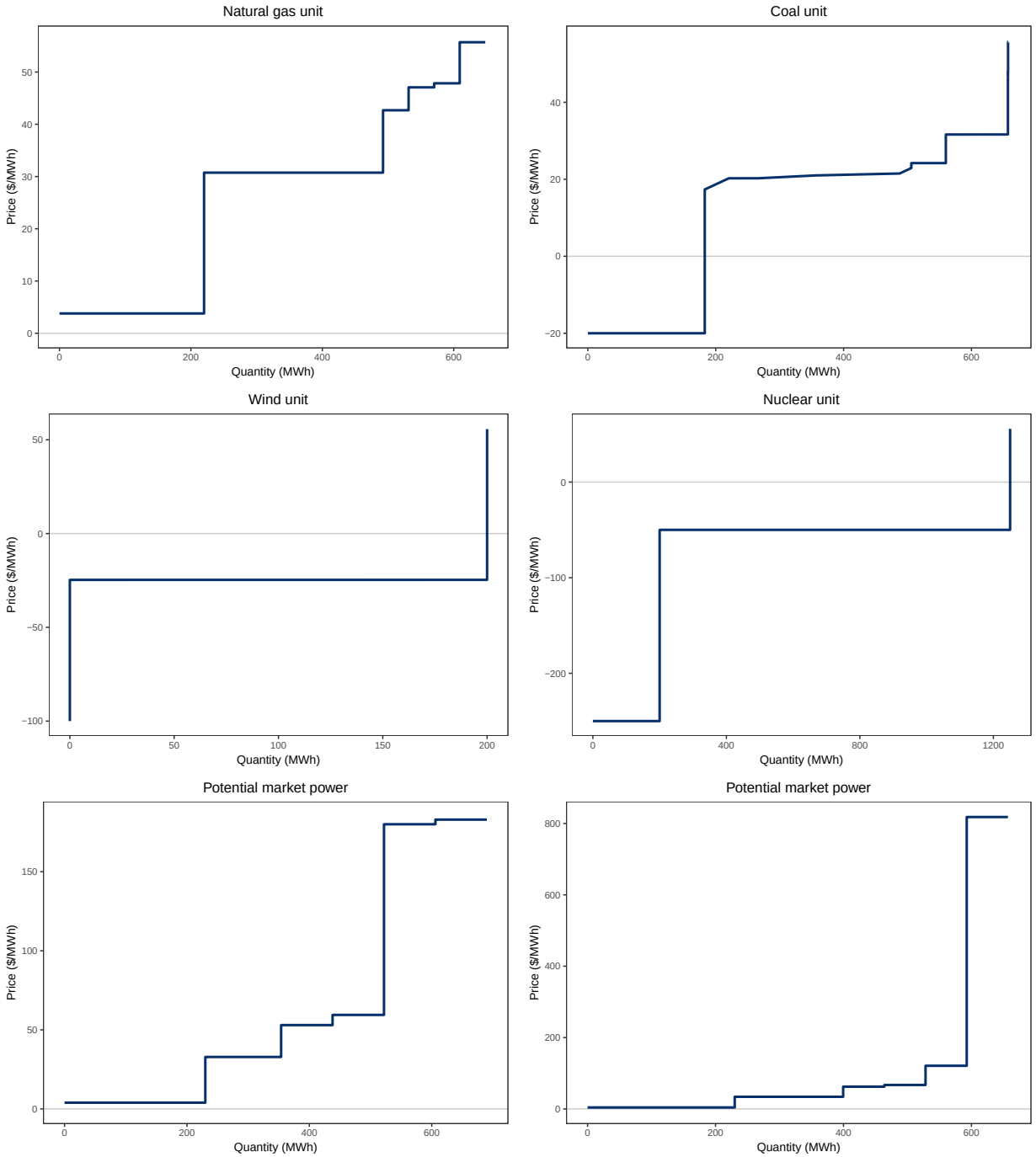
Notes: Both panels map the locational marginal price (LMP) at each unit in the Texas electricity market, but at two different times. The top panel depicts a fully integrated market with no binding constraints, which yields a uniform LMP of roughly $40 per MWh. The bottom panel depicts binding constraints that limit the flow of electricity into the southern portion of the grid and cause the full market to split into two local markets; LMPs in the larger local market remain at roughly $40 per MWh, while LMPs in the smaller local market exceed $3,500 per MWh.

Figure 1.3: Units by fuel type



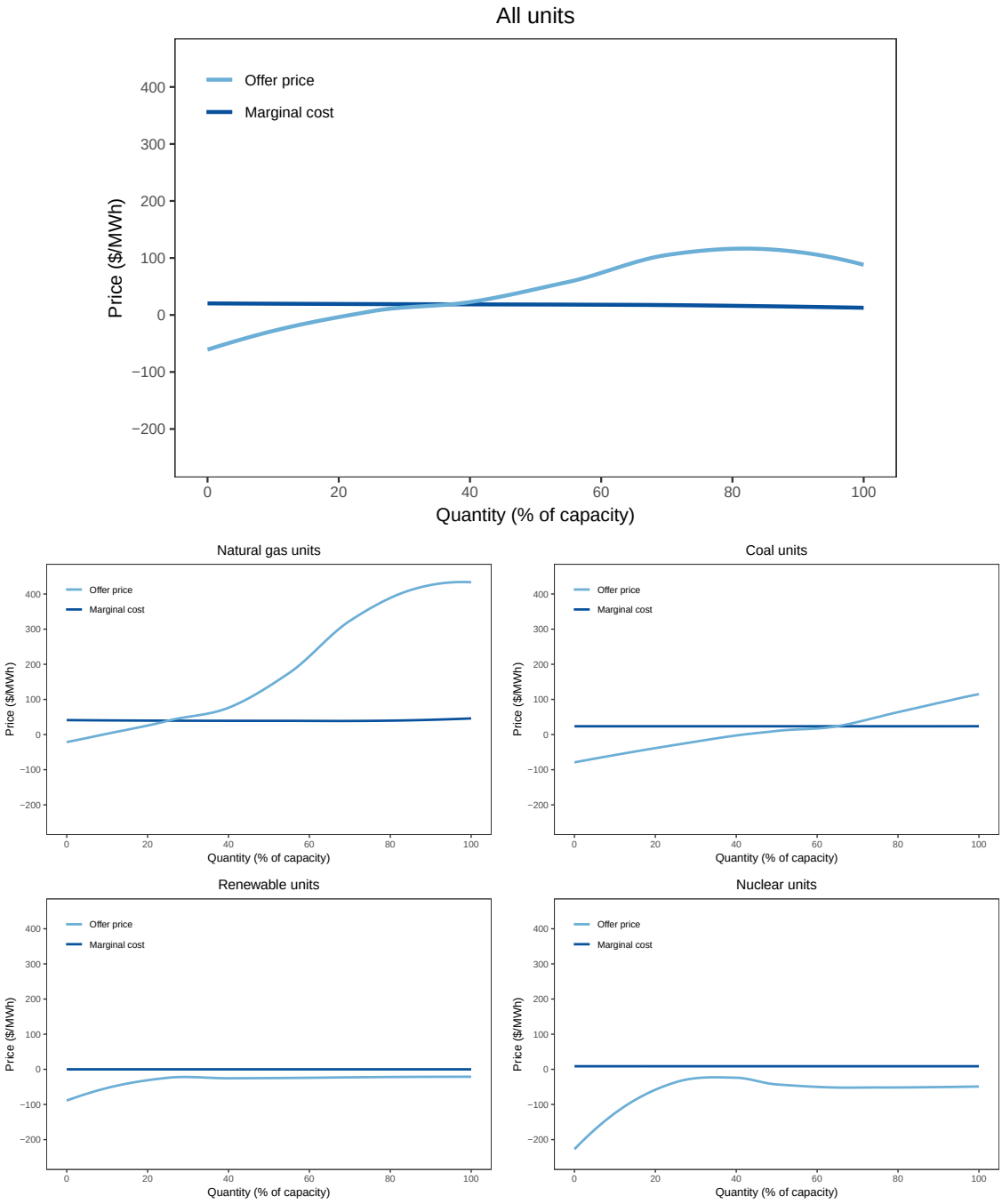
Notes: This figure maps all the power plants that supply power to the Texas electricity market. Each point represents one unit in my analysis. The fuel type of each unit is denoted by the color of the point. Plants are generally spread throughout the state, other than renewable plants, which cluster in areas with the greatest wind resources.

Figure 1.4: Example supply offer curves



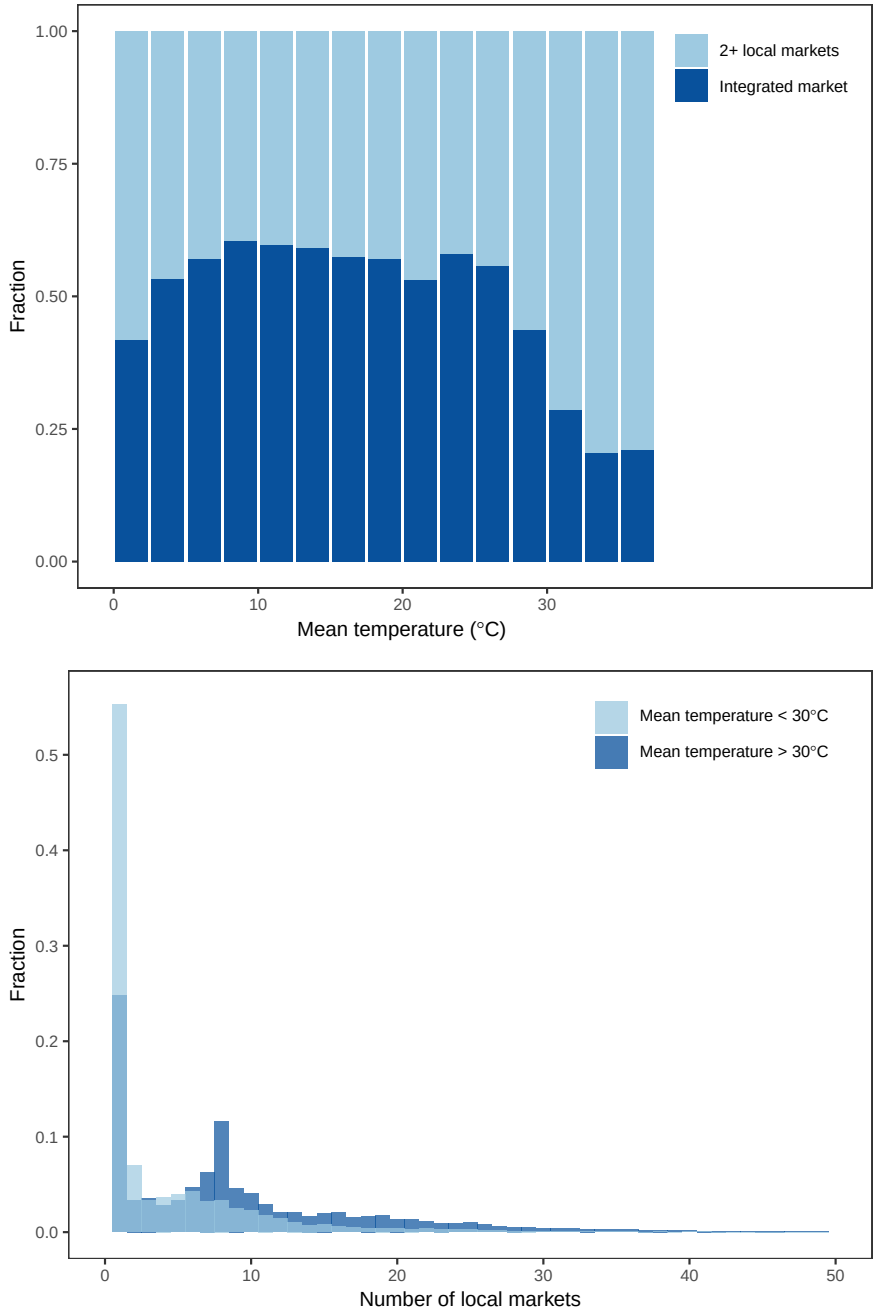
Notes: All six panels plot example offer curves. The top four panels show examples for a natural gas-fired unit, a coal-fired unit, a wind-powered unit, and a nuclear unit, as denoted above each panel. The natural gas-fired example is the same as in Figure 1.1. The bottom two panels depict examples of potential market power, in which the highest offer price greatly exceeds the unit's marginal cost. Note that the axes differ in all panels.

Figure 1.5: Average supply offer curves and marginal costs



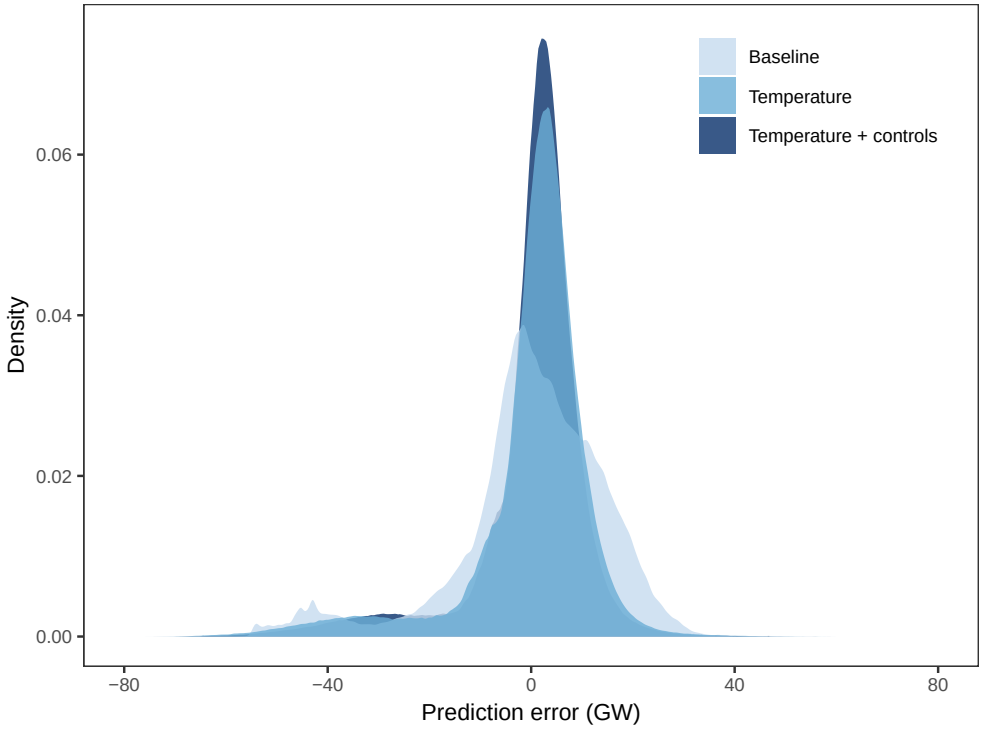
Notes: The top panel depicts the average unit-level offer curve and marginal cost. To construct the average offer curve, I first normalize each offer curve by the unit's nameplate capacity, which yields offer price as a function of the percent of capacity. I then perform a loess regression of offer price on percent of capacity to generate an average offer price at every quantity. Note that not all units offer their full capacity, so the sample changes slightly as quantity increases, particularly as quantity exceeds 90%, which produces the downward slope at large quantities. The bottom four panels depict the average offer curve and marginal cost for each of the four main fuel types, as denoted above each panel.

Figure 1.6: Number of local markets by temperature



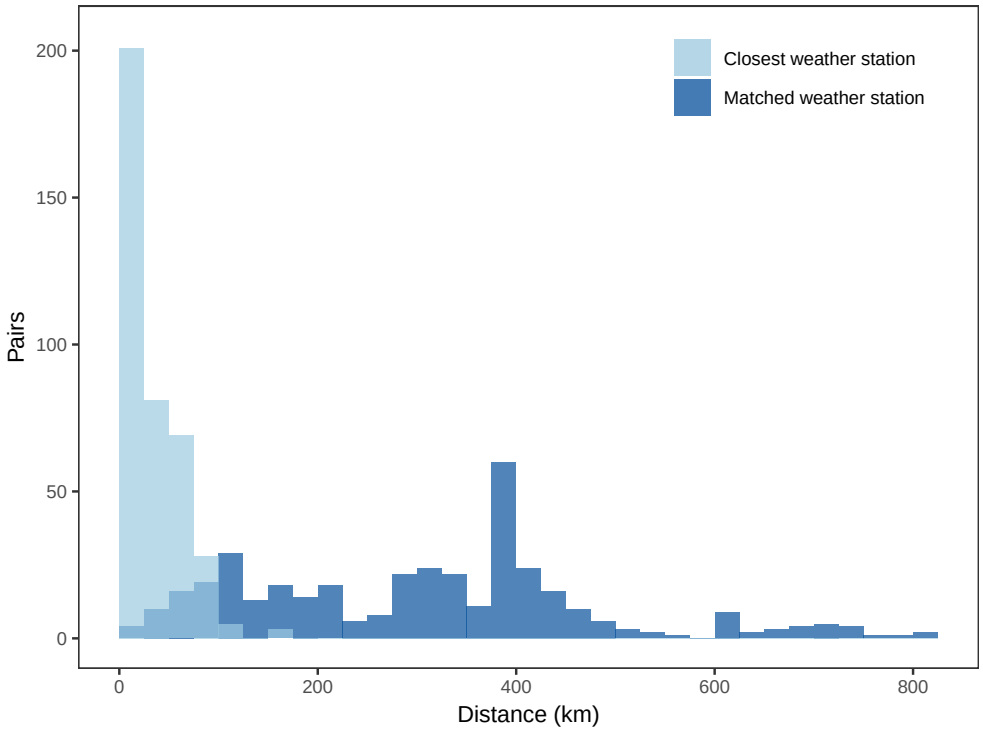
Notes: The top panel plots the fraction of time that the Texas electricity market is a fully integrated market and the fraction of time it is split into multiple local markets, as a function of the mean temperature in the state. Mean temperature is the simple mean of all 75 weather stations in the state. Results are shown by 2.5°C (4.5°F) temperature bins. The first and last bins also include all temperatures below 0°C (32°F) and above 37.5°C (99.5°F), respectively. The bottom panel plots histograms of the number of local markets when the mean temperature is below 30°C (86°F) and when the mean is above this temperature; I omit from this figure the presence of more than 50 local markets, which occurs rarely. At higher temperatures, transmission congestion is more prevalent and the market is more likely to split into many local markets.

Figure 1.7: Distribution of prediction errors



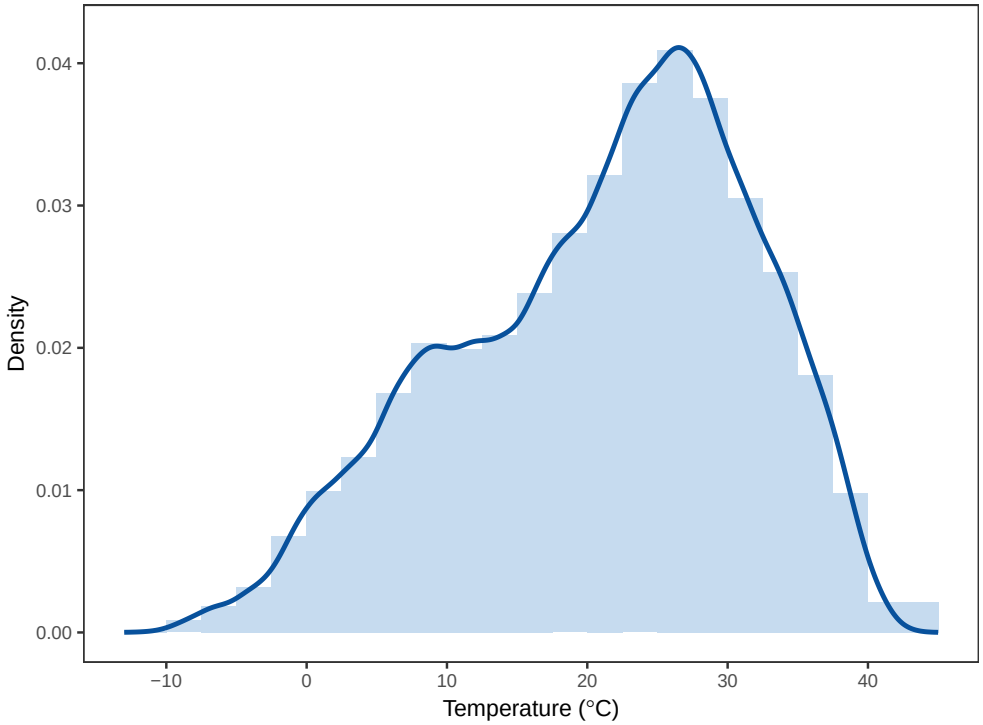
Notes: This figure plots the distributions of prediction errors from three prediction models. Baseline is estimated by Equation (1.7) and represents the raw variation in the data; this model corresponds to Column (1) of Table 1.2. Temperature is estimated by Equation (1.8) with no variables beyond temperature; this model corresponds to Column (2) of Table 1.2. Temperature + controls is also estimated by Equation (1.8) but adds system-wide demand and lagged local market capacity, up to a third-order term for each variable; this model corresponds to Column (3) of Table 1.2. Prediction error is substantially reduced in the Temperature model and further reduced in the Temperature + controls model.

Figure 1.8: Distance from power plants to weather stations



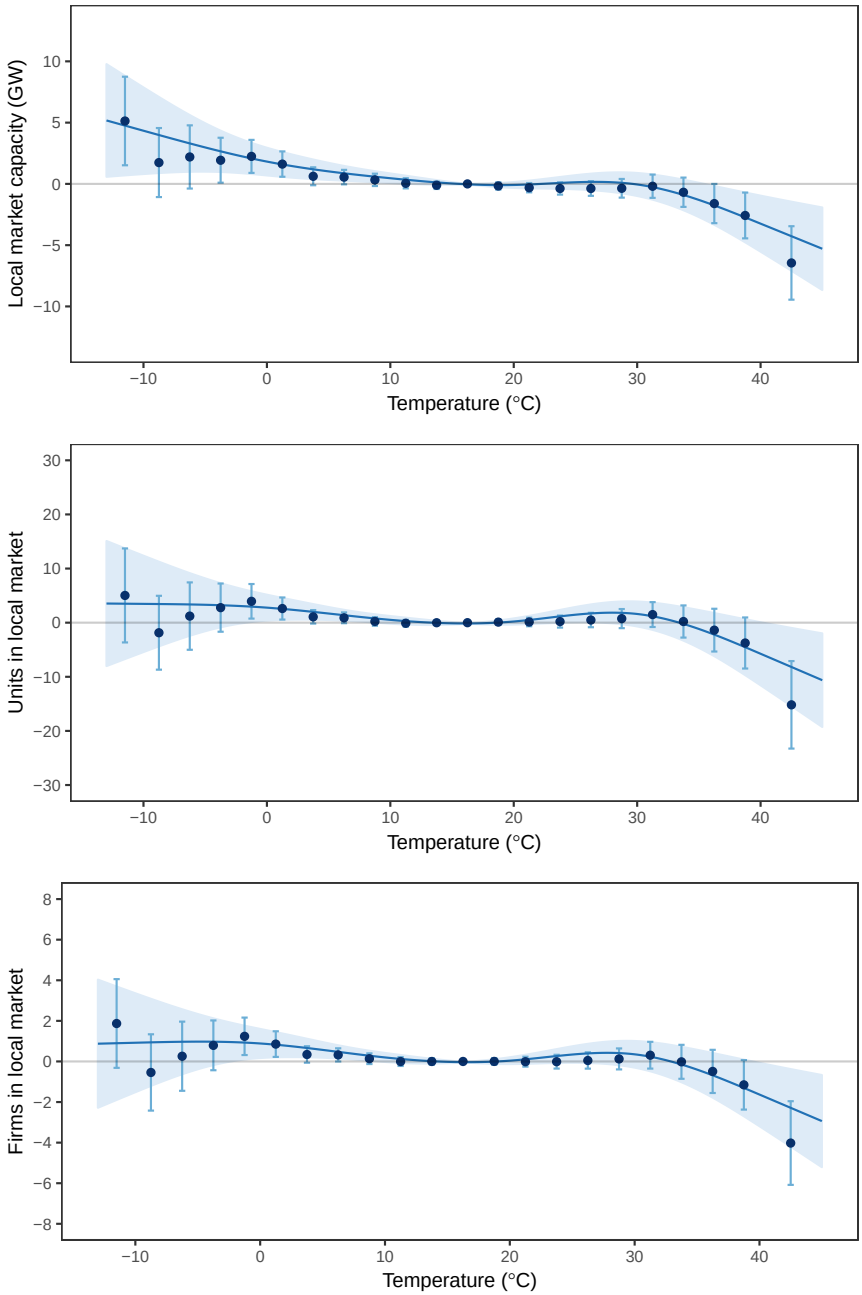
Notes: This figure plots histograms of distances from power plants to weather stations. One histogram corresponds to the weather station closest to the power plant. The other histogram corresponds to the weather station matched to the power plant. The majority of power plants have a weather station within 25 km (15.5 miles), but most are matched to weather stations much farther away.

Figure 1.9: Distribution of matched temperatures



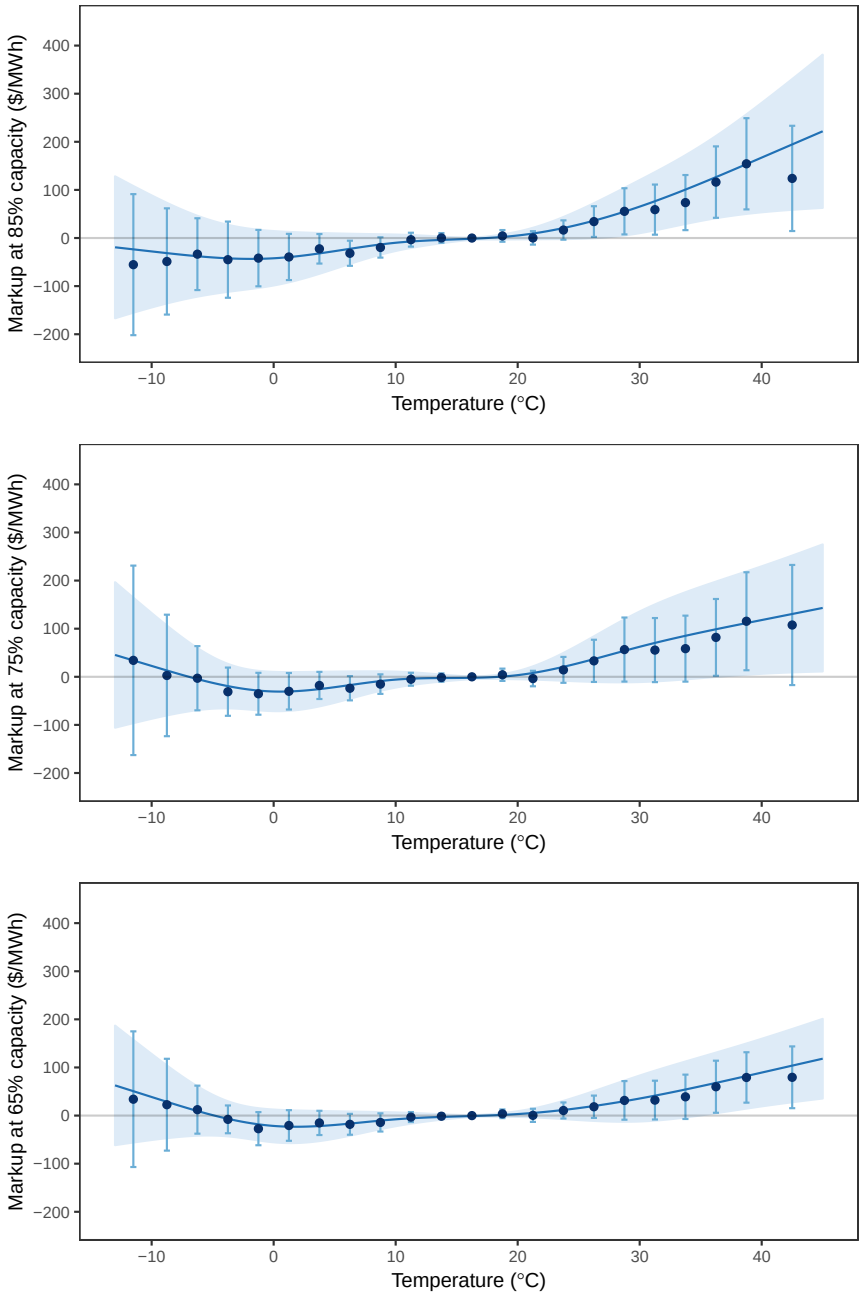
Notes: This figure plots the distribution of matched temperature as both a kernel density and as a histogram of temperature bins. Temperature bins are 2.5°C (4.5°F). The median temperature is 22.7°C (72.9°F), and the modal temperature bin is just above 25°C (77°F). At the upper end of the temperature distribution, temperatures greater than 30°C (86°F) occur in 21.5% of observations, and temperatures exceed 35°C (95°F) in 7.5% of observations.

Figure 1.10: Temperature-response functions for market size measures



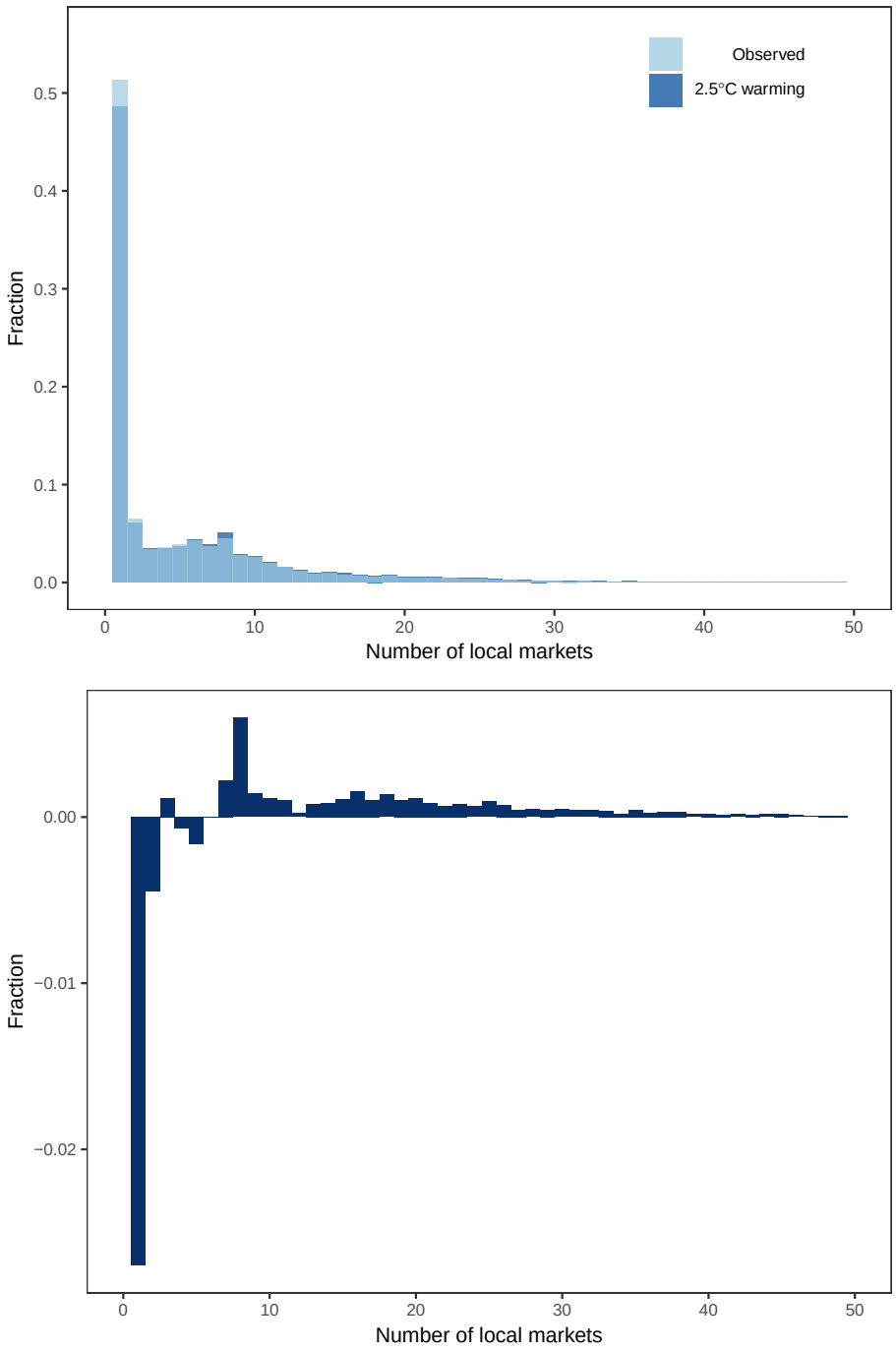
Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (1.11) for a different measure of local market size. From top to bottom, the measures are local market capacity, number of units in the local market, and number of firms in the local market, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure 1.11: Temperature-response functions for markups



Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (1.14) for markup at a different quantity of production. From top to bottom, the quantities are 85%, 75%, and 65%, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure 1.12: Number of local markets under climate change



Notes: The top panel plots histograms of the number of local markets that I observe in the data and that I project to occur under 2.5°C (4.5°F) warming. The bottom panel plots the differences between the two histograms. Negative values indicate a number of local markets that is less likely to occur with warming, and positive values indicate a number of local markets that is more likely. I omit from this figure the presence of more than 50 local markets, which occurs rarely. In the warming scenario, it is less likely for the market to be fully integrated and more likely for the market to split into seven or more local markets.

Tables

Table 1.1: Summary statistics by unit type and firm

	Capacity		Mean generation		
	GW	%	GWh	%	Units
Unit types					
Natural gas	58.9	61.6	15.8	41.8	232
Coal	19.9	20.8	13.8	36.7	32
Renewables	10.7	11.2	3.6	9.6	89
Nuclear	5.1	5.4	4.4	11.8	4
Other	1.0	1.0	0.1	0.2	33
Total	95.6	100	37.7	100	390
Firms					
Luminant	14.5	13.7	8.3	20.1	39
NRG Energy	13.9	13.1	6.3	15.1	44
Calpine	10.1	9.5	4.1	10.0	18
CPS Energy	5.8	5.4	2.2	5.3	19
NextEra Energy	5.3	5.0	2.4	5.8	21
GDF Suez	5.1	4.8	2.1	5.0	15
Lower Colorado River Authority	4.0	3.8	1.6	4.0	26
Exelon	3.9	3.7	0.6	1.5	16
Other (65 firms)	35.2	33.2	10.7	26.0	192
Total	106.1	100	41.3	100	390

Notes: The top panel of this table displays capacity, generation, and number of units by unit type. 99.0% of capacity and 99.8% of generation are accounted for by four fuel types: natural gas, coal, renewables, and nuclear. The bottom panel of this table displays capacity, generation, and number of units by firm. The eight largest firms account for 66.8% of capacity and 74.0% generation, with many small firms. By traditional metrics, this market appears competitive.

Table 1.2: Summary of prediction models

	Baseline	Subset selection		LASSO	
	(1)	(2)	(3)	(4)	(5)
Mean squared error	218.4	149.3	124.0	155.9	129.2
	[112.4]	[105.2]	[83.6]	[107.9]	[85.6]
Median absolute deviation of errors	7.4	4.3	3.8	4.5	4.0
	[0.8]	[1.2]	[1.1]	[1.2]	[1.0]
Interquartile range of errors	15.4	9.3	8.2	9.5	8.5
	[3.0]	[3.9]	[3.4]	[3.7]	[3.3]
Interdecile range of errors	35.9	24.3	21.7	25.6	22.9
	[12.8]	[12.8]	[11.3]	[13.0]	[11.5]
Single, non-parametric temperature		X	X		
Interacted and higher-order temperatures				X	X
Additive explanatory variables			X		
Interacted explanatory variables					X
Hour × month indicators		X	X	X	X

Notes: This table displays results from prediction models. Results are calculated at the unit level, and this table displays the mean and standard deviation of those results. Column (1) is the baseline model estimated by Equation (1.7). Columns (2) and (3) are subset selection models estimated by Equation (1.8). Columns (4) and (5) are LASSO models estimated by Equation (1.9). The subset selection models perform better across all measures than the comparable LASSO models.

Table 1.3: Local market capacity — Effect on markups

	Markup at 85% capacity ($/MWh)			Markup at 75% capacity ($/MWh)			Markup at 65% capacity ($/MWh)		
	OLS (1)	2SLS (2)	2SLS (3)	OLS (4)	2SLS (5)	2SLS (6)	OLS (7)	2SLS (8)	2SLS (9)
Local market capacity (GW)	−1.09	−18.35**	−25.42**	−0.97	−11.98**	−15.24**	−0.72	−9.64***	−13.00***
	(1.02)	(7.63)	(11.22)	(0.97)	(5.35)	(6.46)	(0.79)	(3.05)	(5.03)
First-stage specification		Bins	Spline		Bins	Spline		Bins	Spline
First-stage F-statistic		2.93	6.47		2.75	5.88		2.68	5.28
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating Equation (1.10) with local market capacity as the independent variable. The dependent variable is markup at three different quantities of production, as shown in the column headers. This table reports estimates and standard errors for β. Columns (1), (4), and (7) report results from an OLS regression. The remaining columns report results from 2SLS with the first stage given by Equation (1.11); these regressions use either temperature bins or a natural cubic spline in the first stage, as indicated in the table, and the first-stage results are displayed in the top panel of Figure 1.10. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.4: Number of units in local market — Effect on markups

	Markup at 85% capacity ($/MWh)			Markup at 75% capacity ($/MWh)			Markup at 65% capacity ($/MWh)		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Units in local market	−0.17	−5.74**	−10.27**	−0.19	−3.54**	−5.40**	−0.15	−3.05***	−5.30**
	(0.36)	(2.79)	(5.06)	(0.31)	(1.76)	(2.58)	(0.25)	(0.97)	(2.21)
First-stage specification		Bins	Spline		Bins	Spline		Bins	Spline
First-stage F-statistic		2.11	4.98		2.04	4.65		1.97	4.29
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating Equation (1.10) with the number of units in the local market as the independent variable. The dependent variable is markup at three different quantities of production, as shown in the column headers. This table reports estimates and standard errors for β. Columns (1), (4), and (7) report results from an OLS regression. The remaining columns report results from 2SLS with the first stage given by Equation (1.11); these regressions use either temperature bins or a natural cubic spline in the first stage, as indicated in the table, and the first-stage results are displayed in the middle panel of Figure 1.10. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.5: Number of firms in local market — Effect on markups

	Markup at 85% capacity ($/MWh)			Markup at 75% capacity ($/MWh)			Markup at 65% capacity ($/MWh)		
	OLS	2SLS	2SLS	OLS	2SLS	2SLS	OLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Firms in local market	−0.81	−23.82**	−41.10**	−0.82	−15.26**	−22.95**	−0.65	−12.94***	−21.64**
	(1.31)	(10.64)	(19.17)	(1.16)	(7.20)	(10.25)	(0.95)	(3.91)	(8.47)
First-stage specification		Bins	Spline		Bins	Spline		Bins	Spline
First-stage F-statistic		2.63	5.44		2.51	5.20		2.43	4.87
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating Equation (1.10) with the number of firms in the local market as the independent variable. The dependent variable is markup at three different quantities of production, as shown in the column headers. This table reports estimates and standard errors for β. Columns (1), (4), and (7) report results from an OLS regression. The remaining columns report results from 2SLS with the first stage given by Equation (1.11); these regressions use either temperature bins or a natural cubic spline in the first stage, as indicated in the table, and the first-stage results are displayed in the bottom panel of Figure 1.10. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.6: Local market capacity by owner and unit type — Effect on markups

	Markup at 85% capacity ($/MWh)		
	(1)	(2)	(3)
Local market capacity (GW)	−25.42**		
	(11.22)		
Capacity owned by competing firms (GW)		−50.10**	
		(23.45)	
Capacity owned by own firm (GW)		456.92*	
		(252.19)	
Dispatchable capacity (GW)			−27.59**
			(12.15)
Non-dispatchable capacity (GW)			−2.00
			(73.96)
First-stage specification	Spline	Spline	Spline
First-stage F-statistic	6.47	6.48, 4.92	6.87, 4.57
Local temperature	X	X	X
Unit × hour × month FEs	X	X	X
Hour of sample FEs	X	X	X
Observations	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with measures of local market capacity as the independent variables. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.3. Column (2) reports results with local market capacity split by ownership. Column (3) reports results with local market capacity split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for Columns (2) and (3) are in Appendix A.4. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.7: Robustness to local controls — Effect of local market capacity on markups

	Markup at 85% capacity ($/MWh)			
	(1)	(2)	(3)	(4)
Local market capacity (GW)	−25.42**	−20.69**	−23.70**	−29.80**
	(11.22)	(8.30)	(9.68)	(14.10)
First-stage specification	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	9.33	9.29	5.86
Local temperature	X			X
Local capacity factor			X	X
Unit × hour × month FEs	X	X	X	X
Hour of sample FEs	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with local market capacity as the independent variable. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.3. The remaining columns report results with different sets of local market controls, as denoted in the table. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (2)–(4) are in Appendix A.4. All regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.8: Robustness to matching distance — Effect of local market capacity on markups

	Markup at 85% capacity ($/MWh)			
	(1)	(2)	(3)	(4)
Local market capacity (GW)	−25.42**	−33.32**	−30.32***	−22.24**
	(11.22)	(13.40)	(11.21)	(8.76)
First-stage specification	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	6.38	9.54	11.75
Minimum matching distance (km)	0	100	200	300
Local temperature	X	X	X	X
Unit × hour × month FEs	X	X	X	X
Hour of sample FEs	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with local market capacity as the independent variable. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.3. The remaining columns report results with different matched weather stations. Each unit is matched to the weather station that best predicts the unit's local market capacity and meets the minimum matching distance, as denoted in the table. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (2)–(4) are in Appendix A.4. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 1.9: Temperature selection — Effects on markups

	Markup at 85% capacity ($/MWh)					
	Matched temperature			Local temperature		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−25.42**			−6.23		
	(11.22)			(5.26)		
Units in local market		−10.27**			−2.63	
		(5.06)			(2.24)	
Firms in local market			−41.10**			−8.25
			(19.17)			(7.57)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	4.98	5.44	20.09	19.04	19.29
Local temperature	X	X	X			
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variable is markup at 85% capacity. Columns (1), (2), and (3) duplicate Column (3) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results using local temperature rather than matched temperature. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (4)–(6) are in Appendix A.4. Regressions reported in columns (1)–(3) control flexibly for local temperature; all regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Chapter 2

Spatial Externalities in Groundwater Extraction: Evidence from California Agriculture[1]

2.1 Introduction

Groundwater is a classic example of a common-pool resource (Ostrom 1990). When agent i makes a private decision to extract water from an underground aquifer, this increases the depth of the groundwater below the surface. If many agents share the same aquifer, then agent i's decision to extract imposes an externality on i's neighbors—who must now expend more energy to pump groundwater from greater depths. In theory, a social planner could increase welfare by taxing agent i's marginal unit of groundwater extraction, where the size of the tax would equal the marginal decrease in economic surplus summed across all other agents.

In this paper, we investigate the size of this pumping cost externality for California farmers. California produces 17 percent of U.S. crop value, and farmers rely heavily on groundwater for irrigation. Since these farmers extract water from a few large aquifers, an individual farmer's extraction may increase pumping costs for thousands of neighboring farmers. Moreover, groundwater extraction is largely unregulated in California, and most farmers face no meaningful restrictions on the intensive margin of pumping. Hence, California farmers have little incentive to internalize this open-access externality, whereby their own marginal pumping decisions reduce their neighbors' economic surplus from groundwater consumption.

To estimate the size of this externality, we first must estimate the price elasticity of demand for agricultural groundwater. This has proven difficult, in part because groundwater

[1]This chapter is coauthored with Fiona Burlig and Louis Preonas. We thank Amy Ando, Maximilian Auffhammer, Kathy Baylis, Michael Greenstone, Koichiro Ito, Katrina Jessoe, Dave McLaughlin, Andrew Stevens, Catherine Wolfram, and seminar participants at UC Energy Camp, Mississippi State, University of California Davis, University of Illinois, and the ASSA Annual Meeting for helpful comments and suggestions. Chen Sui, Yixin Sun, and Xinyi Wang provided excellent research assistance. Karen Notsund provided invaluable support in obtaining data.

use is typically neither priced nor measured. However, we leverage the fact that electricity is the main variable input in groundwater extraction. Given data on electricity prices and quantities, along with pump-specific mappings from energy input to groundwater output, we are able to construct accurate measures of groundwater prices and quantities. We assemble a novel dataset that combines (i) confidential electricity consumption data for all agricultural customers served by Pacific Gas & Electric (PGE), California's largest electric utility; (ii) technical pump efficiency audits for nearly 12,000 groundwater extraction points; and (iii) publicly available groundwater measurements across space and time, for all major California aquifers. Using exogenous variation both in PGE's electricity tariff schedules and in average groundwater levels, we identify demand elasticities for both electricity and groundwater.

We estimate farmers' price elasticity of demand for electricity to be -1.17, which is much more elastic than prior estimates of electricity demand in the residential and commercial/industrial sectors. Next, we estimate farmers' price elasticity of demand for groundwater, where we separately identify the effect of groundwater price changes coming from variation in electricity prices vs. variation in groundwater depths. We recover nearly identical groundwater demand elasticities: -1.39 for electricity-induced price changes vs. -1.37 for depth-induced price changes. These statistically indistinguishable estimates suggest that farmers are equally attentive and responsive to either source of variation in groundwater pumping costs, consistent with the predictions of standard Neoclassical theory. We also estimate a single elasticity of demand for groundwater of -1.12, identified using *only* changes in PGE's agricultural electricity tariffs. These estimates are again much more elastic than previous groundwater demand estimates from the existing literature, likely due to farmers' ability to substitute between groundwater and surface water.[2]

Armed with these groundwater demand estimates, we quantify the spatial pumping cost externality. First, we parameterize constant elasticity demand curves for every farm in our sample, in order to calculate farm i's loss in surplus from consuming 1 acre-foot less. We then calculate the marginal increase in groundwater levels for all neighboring farms, resulting from farm i having extracted less.[3] Translating groundwater level increases into pumping cost decreases, we sum the marginal surplus gained by each of farm i's neighbors—equivalent to farm i's short-run pumping cost externality. Our preliminary results suggest that the magnitude of this externality is relatively small (roughly \$2 per acre-foot on average) relative to these farms' pumping costs (averaging \$40 per acre-foot in our sample). We also find substantial variation both within and between California's groundwater basins. In fact, many farms in our sample likely contribute no deadweight loss, as their own marginal surplus loss is greater than the sum of their neighbors' marginal surplus gains. Our results suggest that a relatively small tax on groundwater extraction has the potential to eliminate millions of dollars of deadweight loss annually for California farmers.

This paper makes three main contributions. First, we provide the first large-scale estimates of electricity demand in a major energy-using sector: agriculture in California, one of the most important agricultural sectors in the world. While many studies estimate the

[2]In ongoing work, we are incorporating data on farmers' surface water availability in order to estimate the elasticity of substitution between water sources.

[3] In this draft, we simply calculate effects in concentric circles around each extraction point. Future versions of this analysis will incorporate more realistic and sophisticated hydrogeological assumptions.

relationship between electricity prices and consumption in the residential sector (Alberini and Filippini 2011; Fell, Li, and Paul 2014; Ito 2014; Deryugina, MacKay, and Reif 2017), far fewer have focused on commercial/industrial electricity consumption (Paul, Myers, and Palmer 2009; Jessoe and Rapson 2015; Blonz 2016). To the best of our knowledge, there exists no comparable study estimating the price elasticity of electricity demand in the agricultural sector. By leveraging microdata for thousands of agricultural consumers across PGE's service territory, along with plausibly exogenous changes in farmers' marginal electricity prices, we identify California farmers as relatively elastic electricity consumers.

Second, we estimate the elasticity of groundwater demand for California farmers—a policy-relevant elasticity that has proven elusive due to both data and identification challenges (Mieno and Brozovic 2017). Our empirical strategy overcomes many of these challenges by combining comprehensive electricity consumption data with technical audits of groundwater pumps, and by leveraging exogenous variation in electricity prices (a major component of pumping costs) to credibly identify changes in farmers' effective price of groundwater. Many previous studies have estimated water demand outside the agricultural sector (Hewitt and Hanemann 1995; Renwick and Green 2000; Olmstead, Hanemann, and Stavins 2007), while others have focused specifically on groundwater demand in agriculture (Hendricks and Peterson 2012; Pfeiffer and Lin 2014; Badiani and Jessoe 2015). We provide well-identified groundwater demand estimates, for thousands of farms from one the most important agricultural regions in the world: California's Central Valley.

Third, we compute the pumping cost externality for each farm in our sample. Because groundwater is an exhaustible common-pool resource, it is associated with two first-order externalities: a "stock externality", as farmers do not internalize the full continuation value of the resource and extract too quickly, resulting in a "race to the bottom"; and a "pumping cost externality", as farmers do not internalize how their own extraction lowers groundwater levels, thereby increasing the costs of extraction for neighboring farmers. We focus exclusively on the latter, which is often highlighted in theoretical work (Ostrom 1990; Provencher and Burt 1993) but seldom estimated empirically. Leveraging quasi-experimental variation in electricity prices, our preliminary calculations suggest that the pumping cost externality is likely small relative to California farmers' private costs of extracting groundwater. However, because groundwater demand is elastic, a small corrective tax could yield substantial welfare gains.

This paper proceeds as follows. Section 2.2 provides background on groundwater pumping, California agriculture, and energy use in farming. Sections 2.3 and 2.4 describe our data and empirical strategy. Section 2.5 presents our demand elasticity estimates, which we translate into deadweight loss calculations in Section 2.6. Section 2.7 concludes.

2.2 Background

2.2.1 Groundwater as an open-access resource

Groundwater is a common-pool resource that is typically open access, since most underground aquifers are too large to exclude consumers in a cost-effective manner. This creates

two main externalities from groundwater extraction, commonly known as the "stock externality" and the "pumping cost externality" (Provencher and Burt 1993). The stock externality reflects the finite nature of nonrenewable groundwater stocks. When farmer i removes groundwater from an aquifer in period t, this constrains all users in period $t+1$; because i is not the residual claimant on the full continuation value of the stock, farmers collectively extract faster than the Social Planner's optimal extraction path. While the stock externality is likely important in practice, we lack the empirical machinery to address it in this paper.

We focus exclusively on the pumping cost externality: when farmer i removes a marginal acre-foot of groundwater from an aquifer, the water level falls, thereby increasing pumping costs for other users. While farmer i will internalize the impact of his *own* extraction on his own costs of future extraction, he lacks the incentive to internalize the impacts of his extraction on his *neighbors'* costs. This paper aims to quantify the magnitude of this pumping cost externality—and the associated corrective tax—in the context of California agriculture.

2.2.2 Agriculture in California

California is a key player in global agricultural production, and represented 17 percent of total U.S. crop value in 2016 (Johnson and Cody 2015). California's 77,000 farmers produce over 400 commodities, including close to half of all fruits, nuts, and vegetables grown in the United States. In fact, California is the sole domestic producer of many high-value crops, including almonds, artichokes, olives, and walnuts (California Department of Food and Agriculture 2011).

Water is an essential input for California's agricultural production. Nearly 80 percent of the state's annual water consumption occurs in the agricultural sector, where crop irrigation is the primary end use. California has nearly 8.3 million harvested acres of cropland, 7.9 million of which were irrigated in 2013 (Johnson and Cody 2015). Farmers may have access to surface water and/or groundwater, and each water source is governed by a complex system of rights.[4] California's surface water rights are quite heterogenous across locations and individual users, and they tend to be strictly enforced: farmers using more surface water than allowed by their rights are subject to severe fines.

By contrast, agricultural groundwater rights tend to be far more vague. The typical groundwater right in California is "overlying," meaning that landowners whose property sits above an aquifer have the right to extract the underlying groundwater.[5] The vast majority of groundwater use is unmetered, and users face no variable costs of extraction beyond the energy costs of pumping (Bruno and Jessoe 2018). Hence, a farmer may extract as much groundwater as he chooses, conditional on owning the overlying property rights. During the 2014 drought, California lawmakers passed the Sustainable Groundwater Management Act. This major piece of legislation represents the first statewide regulatory scheme to mitigate overextraction of groundwater, via a series of local groundwater sustainability plans.

[4] See Sawyers (2007) for more details.

[5] There are also "appropriative" groundwater rights, for users who do not own land above the aquifer, but these rights are lower-priority than the overlying rights. Appropriative rights holders may only exercise these rights in the case of a surplus.

However, these plans will not be finalized until 2020 at the earliest, leaving farmers free to extract at will.

2.2.3 Electricity for Pumping

Electricity is an essential input to groundwater pumping. The California Energy Commission reports that water use accounts for 19 percent of California's electricity consumption, and close to 8 percent of the state's energy is used on farms (California Energy Commission 2005). The state's investor-owned utilities spend nearly $50 million annually to improve energy efficiency in the agricultural sector. This makes water use, and agricultural water use in particular, a key component of California's energy policy goals.

To estimate groundwater demand, we exploit the fact that electricity is a major determinant of pumping costs. Several previous papers have used variation in energy costs to estimate the price elasticity of groundwater demand (Hendricks and Peterson 2012; Pfeiffer and Lin 2014; Badiani and Jessoe 2015; Mieno and Brozovic 2017). However, Mieno and Brozovic (2017) point out that these estimates may exhibit bias due to (non-classical) measurement error and/or poor identification. Furthermore, data limitations have restricted previous studies to relatively narrow geographies, which may reduce their external validity. By contrast, our sample includes thousands of farms throughout California's Central Valley, one of the most productive agricultural regions in the world. Our data also allow us to overcome the standard measurement issues and identification challenges—via detailed technical audits that precisely characterize the electricity-to-groundwater conversion factor, and via exogenous variation in Pacific Gas & Electric's (PGE) agricultural electricity tariffs.

2.3 Data

2.3.1 Electricity Data

We begin by estimating how farmers' electricity consumption responds to changes in electricity price. We use confidential customer-level electricity datasets, which PGE's data management team prepared for us under a non-disclosure agreement. These data comprise the universe of agricultural electricity consumers in PGE's service territory, and we observe each customer's monthly bills at the service account level, from 2008–2017. We aggregate service accounts up to 108,172 unique service points (i.e. the physical location of an electricity meter), allowing us to construct "monthified" panel of electricity consumption (in kWh) at the service point (SP) level.[6] We also observe several key covariates for each service point, including its latitude and longitude; an indicator for accounts with solar panels on net-energy metering (NEM), which we drop from our estimation sample; and an identifier to link service point locations to physical electricity meters.

[6] PGE's monthly bill cycles are customer-specific, and billing periods typically do not end at the end of a calendar month. We "monthify" billed kWh for each SP by splitting/weight-averaging multiple bills in a single calendar month, in order to create a SP by month panel. Most SPs have a single service account at each point in time, but service accounts often turn over a given SP.

PGE's offers 23 distinct agricultural tariffs, and our billing data report the particular tariff associated with each monthly bill. Prices on each tariff update multiple times per year, and historic prices are publicly available, along with information on tariff-specific rules and eligibility criteria. This allows us to construct a 10-year panel of hourly volumetric (marginal) electricity prices, which we collapse to the monthly level by taking an unweighted average across hours. Then, we assign each service point a monthly average marginal electricity price in \$/kWh.

Table 2.1 describes each agricultural tariff in detail. Importantly, PGE classifies all agricultural consumers into 5 disjoint categories, based on their physical capital (i.e. pump size; internal combustion engine) and their type of electricity meter (i.e. conventional vs. smart meters).[7] While farmers may switch tariffs *within* categories, they may not switch tariffs *across* categories. This restriction lets us identify price elasticities of demand by instrumenting for farm i's marginal price with the marginal price of its within-category default tariff. Figure 2.1 plots time series of monthly average marginal prices for 3 of the most common tariffs in our estimation sample. Our identifying variation in electricity prices comes from (i) these time series not moving in parallel, and (ii) across-category tariff switching induced by PGE's large-scale smart meter rollout.

2.3.2 Pump Data

To complement our electricity data, we have rich data on agricultural groundwater pumps collected by PGE's Advanced Pumping Efficiency Program (APEP). These data include the universe of APEP-subsidized pump tests from 2011–2017, and we observe detailed measurements and technical specifications for 21,851 unique tests at 17,107 unique pump locations. Importantly, we also observe identifiers for the electricity meter associated with each pump test. This allows us to match pump tests to electricity service points, thereby isolating a sample of 11,851 service points for which agricultural groundwater pumping is confirmed to be a major end-use.[8] We restrict our empirical analysis to this 11 percent subset of agricultural service points, in order to best isolate groundwater pumpers and avoid incorporating other agricultural electricity end-uses.[9]

Table 2.2 reports summary statistics for this subset of agricultural service points (in the right column). Compared to the full set of PGE's agricultural customers, APEP-matched service points tend to consume nearly twice as much electricity, and tend to pay lower marginal prices. Only 24 percent of service points shift across tariff categories, and the vast majority of switches are triggered by PGE's smart meter rollout. Figure 2.2 reveals that

[7] Conventional meters record electricity consumption using an analog dial, whereas smart meters can digitally store the full time profile of consumption. During our sample period, PGE gradually phased out conventional meters, replacing them with smart meters capable of supporting time-varying electricity pricing.

[8] Pumping is likely the *only* end-use at most matched service points, as PGE typically installs a dedicated meters for each pump.

[9] We are currently working on using satellite images to predict whether service points outside the APEP-matched sample are also groundwater pumps. We hope to incorporate these farms into future analysis, as there are likely many groundwater pumps that never received an APEP-subsidized pump test.

APEP-matched service points are heavily concentrated in California's Central Valley, and appear to be a geographically representative subset of PGE's agricultural customers.

Besides helping us identify a subset of agricultural consumers who pump groundwater, APEP data allow us to characterize pump-specific groundwater production functions. Groundwater extracted is Leontief in electricity (for pumps with electric motors), and 1 kWh of electricity in will yield a particular volume of groundwater out (measured in acre-feet (AF)). This kWh per AF relationship is governed by physics:

$$\frac{\text{kWh}}{\text{AF}} = \text{kW} \div \frac{\text{AF}}{\text{hour}} = \frac{[\text{Lift (feet)}] \times [\text{Flow (gallon/minute)}]}{[\text{Operating pump efficiency (\%)}] \times [\text{Constant}]} \div \frac{\text{AF}}{\text{hour}} \tag{2.1}$$

The power (kW) required to pump 1 acre-foot is directly proportionate to both the vertical distance the water must travel to the surface (i.e. lift) and the speed at which the water travels (i.e. flow). It is inversely proportionate to rate at which the pump converts electric energy into the movement of water (i.e. operating pump efficiency). We can simplify (2.1) by converting from gallons to acre-feet:

$$\Rightarrow \quad \frac{\text{kWh}}{\text{AF}} = \frac{[\text{Lift (feet)}] \times [\text{Constant}]}{\text{Operating pump efficiency (\%)}} \tag{2.2}$$

For each APEP pump test, we observe measurements of kWh/AF, operating pump efficiency, flow, and lift. We also observe the standing water level, or the baseline groundwater depth in the absence of pumping. Because pumping lowers the water level at a given location, standing water levels help us more accurately calibrate how changes in aquifer depth impact lift for each pump.[10]

2.3.3 California Water Data

While a given farm's pumping technology is relatively constant in the short run, its kWh/AF conversion factor is sensitive to short-run changes in groundwater levels. In order to capture these short-run shocks in pumping costs, we use publicly available groundwater data from California's Department of Water Resources collected under the California Statewide Groundwater Elevation Monitoring (CASGEM) Program.[11] These data report depth below the surface at 16,015 unique monitoring stations during our 2008–2017 sample period, with an average of 25 measurements at each location at different points in time. We rasterize all measurements within each month (and quarter), using inverse distance weighting to interpolate a gridded two-dimensional surface of average depth at each point in space. This allows

[10] Lift is (approximately) the sum of the standing water level, drawdown (i.e., how much pump i impacts its own depth), and other pump-specific factors (e.g., discharge pressure, gauge corrections, height of the pump above the surface). Drawdown depends on rate of extraction (i.e. flow) and the physical properties of the substrata. Greater flow increases drawdown, as water levels fall with faster extraction. More transmissive (or porous) rock formations have lower drawdown, because water levels are able to horizontally reequilibrate more quickly.

[11] These data are available at: `https://water.ca.gov/Programs/Groundwater-Management/Groundwater-Elevation-Monitoring--CASGEM`

us to construct a monthly (and quarterly) panel of estimated groundwater depths at each electricity service point.

We assign each service point to both a groundwater basin and a water district, using publicly available shapefiles maintained by the California Department of Water Resources.[12] Groundwater basins are broadly defined by stratigraphic barriers through which water does not travel horizontally. We control for annual changes in water levels that impact all farms within the same water basin. Water districts are administrative entities that govern farmers' annual allocations of surface water. Because groundwater and surface water are obvious substitutes, we control for annual shocks to farms' surface water allocations at the water district level. This helps to isolate variation in pumping behavior driven by variation in pumping costs, rather than variation in the availability of groundwater substitutes.[13]

2.3.4 Groundwater Prices and Quantities

We merge the above data sources into a panel of groundwater prices and quantities, at the service point by month level. To convert from electricity (kWh or \$/kWh) to groundwater (AF or \$/AF), we simply need to populate a kWh/AF conversion factor for every panel observation. We construct estimates for kWh/AF by parameterizing Equation (2.2) using (i) monthly (or quarterly) rasters of groundwater depths at each service point; (ii) pump-specific conversions between standing water level and lift, as calculated from APEP pump tests; and (iii) APEP-measured operating pump efficiencies. We take unweighted averages of APEP variables across multiple pumps within a single service point; we also extrapolate each service point's first pump test backwards, extrapolate its last pump test forwards, and interpolate between multiple pump tests using a triangular kernel in time.

Table 2.3 reports summary statistics for this merged panel dataset. We observe 3.45 unique pump tests for the average APEP-matched service point, and APEP data reveal substantial (cross-sectional) variation in operating pump efficiencies and kWh/AF conversion factors. Our constructed kWh/AF estimates tend to moderate extreme values, which compresses the right tail of *measured* kWh/AF (while also slightly shifting this distribution left). Interestingly, implied marginal groundwater prices exhibit far less seasonal variation than marginal electricity prices. This is because groundwater levels tend to be higher in summer months (compared to winter months), which tends to reduce (constructed) kWh/AF in months when electricity prices are highest.

[12] Water basin shapefiles are available at `https://water.ca.gov/Programs/Groundwater-Management/Bulletin-118`. Water district shapefiles are available at `https://data.cnra.ca.gov/dataset/water-districts`.

[13] We are currently working to incorporate additional spatial data products, including (i) shapefiles of common land units, which correspond to agricultural fields; (ii) shapefiles of parcels, which roughly correspond to farms; and (iii) the USDA Cropland Data Layer, which uses satellite imagery to classify fields by crop type. Together, these data will enable us to aggregate up from service points to farms, and to build a panel of crop choices at the farm-year level.

2.4 Empirical Strategy

This section outlines our empirical strategy for estimating farmers' demand for groundwater pumping. First, we estimate price elasticities of demand for electricity, for the full sample of agricultural consumers where we can match an electricity meter to a groundwater pump. Next, we estimate price elasticities of demand for groundwater, by translating prices/quantities of electricity into prices/quantitites of water using data on (i) technical pumping production functions and (ii) groundwater depths across space and time.

2.4.1 Electricity Demand

We estimate monthly electricity demand using the following specification:

$$\sinh^{-1}\left(Q_{it}^{\text{elec}}\right) = \beta \log\left(P_{it}^{\text{elec}}\right) + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.3}$$

The dependent variable is kWh consumed at electricity service point i in month t, transformed using the inverse hyperbolic sine function (which closely approximates the natural log transformation but includes zero in its support).[14] P_{it}^{elec} is unit i's marginal electricity price (in \$/kWh), averaged across all hours in month t. We include unit by month-of-year fixed effects (γ_i), in order to control for within-pump/month average consumption (e.g., service point i in March). We also include month-of-sample fixed effects (δ_t) to control for trends in both electricity prices (which rise over time) and pumping behavior. Alternative specifications include groundwater basin by year fixed effects (to control for time-varying trends in groundwater depth across basins), water district by year fixed effects (to control for annual shocks to surface water allocations), and unit-specific linear time trends. We two-way cluster standard errors by service point and month-of-sample, which accommodates both arbitrary within-unit serial correlations and arbitrary spatial correlations with each monthly cross-section.

To identify the demand elasticity in Equation (2.3), we must purge any endogenous variation in unit i's marginal electricity price. PGE's agricultural tariff *schedules* are the outcome of statewide regulatory proceedings, and marginal prices are linear in kWh consumed.[15] While an individual farmer cannot plausibly influence how PGE sets prices, most farmers may select between alternative tariff schedules—effectively choosing which marginal electricity price they face. PGE restricts this choice to be within 5 tariff categories defined by farmers' physical capital (e.g. pump size and type) and type of electric meter (i.e. conventional vs. smart meters). We instrument for unit i's marginal price with the marginal price of the default tariff *within* unit i's category (i.e. bolded tariffs in Table 2.1), which eliminates selection bias from a high-volume pumper choosing a tariff with advantageously low volumetric prices.

[14] Since 15 percent of observations in this panel are zeroes, we apply the inverse hyperbolic sine transformation to avoid dropping months where farms consume zero kWh for groundwater pumping.

[15] By contrast, PGE's residential electricity tariffs use increasing block pricing, where a household's marginal price is endogenous to its own consumption (Ito 2014). Linear marginal prices simplify our estimation of agricultural electricity demand, because farm i's marginal price is determined *solely* by its tariff schedule.

Farmers may also shift *across* tariff categories, which could similarly bias our elasticity estimates. If such a shift reflects a change in physical pumping capital—for example, upgrading from a < 35 hp pump to a ≥ 35 hp pump—then a change in marginal price (or within-category marginal default price) would coincide with a mechanical increase in electricity consumption. We control for such endogenous changes in price by interacting unit fixed effects with a categorical variable for the 3 types of physical capital that define PGE tariff categories: small pumps, large pumps, and auxiliary internal combustion engines. On the other hand, if unit i shifted across categories because PGE replaced its conventional meter with a smart meter, we would not expect such a shift to coincide with any other changes to unit i's pumping behavior.[16] Hence, meter-induced shifts in tariff categories are unlikely to lead to *endogenous* changes in unit i marginal electricity price. We also instrument with *lagged* default prices to purge potential endogeneity in the timing of unit i's smart meter installation.

2.4.2 Groundwater Demand

We seek to estimate the causal effect of groundwater price on groundwater consumption, and this demand elasticity is linearly approximated by the coefficient β:

$$\log\left(Q_{it}^{\text{water}}\right) = \beta \log\left(P_{it}^{\text{water}}\right) \tag{2.4}$$

We construct Q_{it}^{water} and P_{it}^{water} using the *estimated* conversion factor $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$, which has measurement error and is also potentially endogenous. Hence, the same measurement error and endogeneity is present on both the left-hand side and the right-hand side of Equation (2.4). We can rewrite this expression decomposing $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$ on both sides:

$$\log\left(Q_{it}^{\text{elec}}\right) - \log\left(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}\right) = \beta\left[\log\left(P_{it}^{\text{elec}}\right) + \log\left(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}\right)\right] \tag{2.5}$$

Rearranging:

$$\log\left(Q_{it}^{\text{elec}}\right) = \beta \log\left(P_{it}^{\text{elec}}\right) + (\beta + 1)\log\left(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}\right) \tag{2.6}$$

This expression is algebraically equivalent to Equation (2.4), but it isolates the endogenous estimated conversion factor in one right-hand-side variable. We estimate an analogous regression specification:

$$\sinh^{-1}\left(Q_{it}^{\text{elec}}\right) = \beta^{\text{e}} \log\left(P_{it}^{\text{elec}}\right) + (\beta^{\text{w}} + 1)\log\left(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}\right) + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.7}$$

[16] During our 2008–2017 sample period, PGE gradually installed smart meters for the vast majority of its customers. The timing of PGE's smart meter rollout was determined by institutional and geographic factors, which were outside of customers' control. Previous research has established that PGE did not specify the rollout to target customers with particular usage patterns (Blonz 2016).

This specification is similar to Equation (2.3), except that we can now interpret β^{e} and β^{w} as the price elasticity of demand for groundwater. We allow this elasticity to vary depending on the source of variation in pumping costs—groundwater depths may be more salient to farmers than electricity prices, or vice versa.[17] As in the electricity regressions, we purge electricity price endogeneity by instrumenting P_{it}^{elec} with within-category default prices (see description above).

To identify β^{w}, we must overcome three potential sources of bias. First, farmers may choose to alter their pumping technologies in order to change $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$, and such changes are likely correlated with Q_{it}^{elec}. Second, $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$ is a function of unit i's groundwater depth, which is mechanically linked to Q_{it}^{elec}—when unit i consumes electricity to extract groundwater, its localized groundwater level falls, thereby increasing $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$. Third, $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$ incorporates measurement error both from interpolating rasterized groundwater depths across space and from interpolating/extrapolating unit i's APEP measurements across time.

We instrument for $\log\left(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}\right)$ using logged groundwater depth averaged across unit i's full groundwater basin.[18] This purges potential endogeneity driven by changes in pumping technologies, and eliminates bias induced by measurement error in unit i's pump specifications in month t. It also breaks the mechanical relationship between $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$ and Q_{it}^{elec}, as farm i's extraction should have a negligible contemporaneous effect on average groundwater levels across the whole basin. Finally, instrumenting with basin-wide average depth mitigates measurement error from having spatially interpolated groundwater measurements into a (potentially overfit) gridded raster.

While Equation (2.7) isolates the shared endogenous component of Q^{water} and P^{water} on the right-hand side, a more standard approach would be to estimate:

$$\sinh^{-1}\left(Q_{it}^{\text{water}}\right) = \beta \log\left(P_{it}^{\text{water}}\right) + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.8}$$

We also estimate Equation (2.8), instrumenting for $\log\left(P_{it}^{\text{water}}\right)$ using the logged average marginal electricity price of unit i's within-category default tariff (i.e., the same instrument from Equation (2.3)). This isolates changes in the effective price of groundwater driven *only* by plausibly exogenous changes in the marginal electricity price. It also removes right-hand-side measurement error in $\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}$, thereby preventing measurement error in Q_{it}^{water} from biasing our point estimates.[19]

[17] A strict Neoclassical interpretation would assume $\beta^{\text{e}} = \beta^{\text{w}}$, as the optimizing farmer should respond to all short-run changes in P_{it}^{water} identically.

[18] We instrument with groundwater depth in logs (rather than levels) because logging both sides of Equation (2.2) implies that $\log\left(\text{kWh}/\text{AF}_{it}\right)$ is linear in $\log\left(\text{lift}\right)$, and a percentage change in depth should yield a similar percentage change in lift.

[19] In most cases, classical measurement error on the left-hand side does not bias point estimates. However, consider the regression $(Y_i + \eta_i) = \beta(X_i + \omega_i) + \varepsilon$, where η_i and ω_i each denote classical measurement error, and where $\text{Cov}(\eta_i, \omega_i) \neq 0$. After conditioning on $(X_i + \omega_i)$, the remaining measurement error on the left-hand side is no longer *conditionally* classical, and could bias $\hat{\beta}$. In Equation (2.8), measurement error

2.5 Results

2.5.1 Electricity Demand

Table 2.4 reports results from estimating Equation 2.3. Column (1) does not instrument for marginal electricity price, yielding an unidentified $\hat{\beta}$ estimate. Column (2) instruments using unit i's within-category default marginal price, which eliminates bias from farmers choosing their own electricity tariffs. The direction of this bias is not obvious *ex ante*, because farmers are choosing between tariffs with both volumetric (\$/kWh) and fixed (\$/kW) price components.[20] Comparing Columns (1) vs. (2), we see that on average, farmers with higher electricity consumption tend to choose tariffs with relatively low fixed charges per kW and relatively high prices per kWh.

Column (3) eliminates the other source of price endogeneity by interacting unit fixed effects with indicators for (i) small pumps ($<$ 35 hp), (ii) large pumps ($\geq$ 35 hp), and (iii) auxiliary internal combustion engines. While only 5 percent of units shift across tariff categories due to changes in their physical capital, the resulting simultaneous changes in Q_{it}^{elec} and P_{it}^{elec} induce substantial bias in Column (2) point estimate. Column (3) reports our preferred estimate of -1.17, after having purged both sources of price endogeneity.

The magnitude of this elasticity estimate is surprisingly large, considering that electricity demand tends to be extremely inelastic in other contexts. The literature on electricity demand has focused heavily on the residential sector, and recent estimates have found elasticities of -0.08 to -0.48 in the short run (Reiss and White 2005; Alberini and Filippini 2011; Fell, Li, and Paul 2014) and -0.09 to -0.73 in the medium-to-long run (Alberini and Filippini 2011; Ito 2014; Deryugina, MacKay, and Reif 2017).[21] Fewer estimates exist for commercial or industrial electricity demand. Paul, Myers, and Palmer (2009) estimate commercial/industrial elasticities of -0.11 to -0.16 in the short run, and -0.29 to -0.40 in the long run. Jessoe and Rapson (2015) find no demand response to dynamic pricing in these sectors, while Blonz (2016) estimates elasticities of -0.08 to -0.22 using hourly price variation for PGE's small commercial/industrial customers. To our knowledge, we provide the first large-scale estimates of electricity demand elasticities in the agricultural sector.

Columns (4)–(6) report three additional elasticity estimates, each intended to assuage any remaining concerns over price endogeneity. Column (4) includes separate year fixed effects for each water basin and each water district, to control for potential time-varying confounders

from $\widehat{\text{kWh}/\text{AF}}_{it}$ enters directly on the right-hand side and inversely on the left-hand side. Instrumenting with default electricity prices neutralizes this correlation between left-hand- vs. right-hand-side measurement error.

[20] PGE tariffs with relatively high volumetric (i.e. marginal) prices tend to have relatively low fixed prices, and vice versa. Two farmers with the same average electricity consumption may optimally choose different tariffs. Suppose farmer A operates a 300 hp pump for 50 hours per month, while farmer B operates 50 hp pump for 300 hours per month. Farmer A should prefer a low fixed price and a high volumetric price, while farmer B should prefer a high fixed price and a low marginal price.

[21] These estimates uses monthly or annual variation in electricity prices, which aligns with our empirical strategy. Other studies leverage hourly variation in electricity prices have estimated electricity demand elasticities ranging from -0.03 to -0.25 (Wolak 2011; Jessoe and Rapson 2014; Fowlie et al. 2017; Ito, Ida, and Tanaka 2018).

related to water depth or surface water availability. The resulting point estimate of -1.02 is similar, albeit slightly attenuated. Column (5) instruments with the 6- and 12-month lags of the default price (rather than the contemporaneous default price), to account for potential endogeneity in the timing of PGE's smart meter rollout.[22] This yields a nearly identical point estimate, implying that farmers' electricity consumption did not meaningfully change in anticipation of a smart meter installation. Finally, Column (6) adds 11,175 unit-specific linear time trends, to confirm that we are not identifying $\hat{\beta}$ solely off of monotonic trends in price and quantity. The resulting point estimate of -0.76 is attenuated, as linear trends remove much of the (good) variation in electricity prices over time. Even so, we still find a tightly estimated elasticity that is substantially larger than virtually all previous estimates for the elasticity of electricity demand.

2.5.2 Groundwater Demand

Table 2.5 presents our results for estimating farmers' groundwater demand. Each column estimates Equation (2.7) using our preferred strategy for identifying the elasticity with respect to the electricity price: instrumenting for $\log(P_{it}^{\text{elec}})$ with within-category default prices, and interacting unit fixed effects with indicators for each category of physical pumping capital. Note that we report $\hat{\beta}^{\text{e}}$ and $\hat{\beta}^{\text{w}}$, where the latter subtracts 1 from the regression coefficient on $\log(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it})$. We interpret each coefficient as the elasticity of demand for groundwater with respect to one component of the price of groundwater, holding the other component constant.

Column (2) reports our preferred estimates of $\hat{\beta}^{\text{e}}$ and $\hat{\beta}^{\text{w}}$, where we instrument for $\log(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it})$ with logged groundwater depth in month t averaged across unit i's groundwater basin. Comparing $\hat{\beta}^{\text{w}}$ in Columns (2) vs. (1), instrumenting with average depth appears to alleviate bias due to measurement error in $\log(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it})$.[23] The exclusion restriction requires that unit i's pumping behavior have no contemporaneous impact on basin-wide average groundwater depths. Such feedback effects between the dependent variable and the instrument would be extremely unlikely for three reasons: (i) unit i is small relative to the geographic footprint of its groundwater basin; (ii) thousands of other pumpers are also extracting from the same basin; (iii) basin-wide average groundwater levels do not instantaneously reequilibrate after extraction at one point in space. Column (3) restricts the sample to the 3 largest groundwater basins, each of which has over 1,000 units in our estimation

[22] Recall that farmers may shift across tariff categories (inducing changes to their within-category default price) due to *either* changes in their physical capital *or* the installation of a smart meter.

[23] We discuss three potential sources of bias in β^{w} in Section 2.4.2 — (i) endogenous changes to pumping technologies, (ii) the mechanical relationship between extraction and depth at a given location, and (iii) measurement error. Bias from (i) and (ii) appear unlikely, as they should bias our β^{w} away from zero, rather than towards zero.

sample.[24] The resulting $\hat{\beta}^{w}$ estimate is quite similar, which should assuage concerns that the instrument is invalid due to a few large farms located in very small groundwater basins.

The magnitudes of our $\hat{\beta}^{e}$ estimates are quite similar to results from the electricity-only regressions, especially comparing $\hat{\beta}^{e} = -1.21$ from Column (1) of Table 2.5 the analogous estimate of $\hat{\beta} = -1.17$ from Column (3) of Table 2.4. This is not surprising, since Equation (2.7) simply adds one regressor to Equation (2.3). However, $\hat{\beta}^{e}$ is surprisingly close to our instrumented $\hat{\beta}^{w}$ estimate (-1.39 vs. -1.37). This implies that a 1 percent change in the effective price of groundwater has the *same* effect on farmers' pumping behavior, whether that change comes via their marginal electricity price or via their pump's kWh/AF conversion factor. It also suggests that farmers are quite attentive to their true costs of pumping, and that they reoptimize their pumping behavior identically in response to either type of price variation—as Neoclassical theory would predict.

Similar to our elasticity estimates for electricity, our groundwater elasticity estimates are also quite large relative to the existing literature. Recent studies have also exploited variation in energy prices, but yielding far smaller magnitudes: Hendricks and Peterson (2012) find an elasticity of -0.10, and Pfeiffer and Lin (2014) find an elasticity of -0.27 (both for agricultural groundwater in Kansas). Bruno and Jessoe (2018) estimate demand elasticities of -0.17 to -0.22 within the Coachella Valley of California, which is a unique setting where groundwater extraction is directly priced. Previous studies have also estimated farmers' elasticity of demand for surface water, most notably Hagerty (2018), who finds an elasticity of -0.23 for surface water in California agriculture. While estimates of surface water demand are often as large as -0.8 in specific locations (Schoengold, Sunding, and Moreno 2014; Hagerty 2018), we find agricultural groundwater demand to be even more elastic *on average*.[25] Substitution between groundwater and surface water is likely a major factor explaining the large magnitudes of our elasticities estimates.

Columns (4)–(6) report three alternate versions of our preferred estimates in Column (2). First, to account for the inherent tradeoff between spatial density vs. temporal frequency of groundwater measurements, Column (4) re-estimates Equation (2.7) using groundwater data rasterized at the quarterly (rather than monthly) level. Whereas our preferred monthly rasters are able to capture groundwater measurements at greater temporal frequency, quarterly rasters have greater accuracy in the cross-section by incorporating more distinct measurement sites. The resulting $\hat{\beta}^{w}$ estimate increases in magnitude, however the average depth instrument has less predictive power at the (coarser) quarterly level. Column (5) includes water basin by year and water district by year fixed effects, yielding only slightly attenuated point estimates despite eliminating much of the variation in the average depth instrument. In Column (6), we instrument with 6- and 12-month lags of average depth (rather than contemporaneous depth), as it is possible (albeit unlikely) that farmers pump less in months with lower groundwater levels for some reason other than pumping costs. These lagged instru-

[24] These basins are the San Joaquin Valley, the Sacramento Valley, and the Salinas Valley. The number of agricultural groundwater pumpers in each basin is likely much larger, as our estimation sample comprises only the subset of PGE customers that we can confident match to an APEP-subsidized pump test.

[25] Estimates for urban water demand have found similar elasticities, ranging from -0.10 to -0.76 (Nataraj 2011; Ito 2013; Baerenklau, Schwabe, and Dinar 2014; Wichman 2014; Buck et al. 2016; Wichman, Taylor, and Haefen 2016; Hagerty 2018).

ments substantially increase $\hat{\beta}^{\text{e}}$ and $\hat{\beta}^{\text{w}}$; however, the small first stage F-statistic indicates a weak instrument, and we interpret these results with caution.

Table 2.6 reports results from estimating Equation (2.8), with groundwater quantity as the dependent variable, and instrumenting for the composite groundwater price with default electricity prices. While these estimates identify $\hat{\beta}$ using *only* variation in default electricity prices, Table 2.5 demonstrates that farmers respond almost identically to variation in *either* component of their effective groundwater price. The resulting point estimates are quite similar to, but slightly smaller than, our electricity demand estimates.[26] Interestingly, combining P_{it}^{water} into a single regressor removes much of the variation used to estimate separate coefficients in Table 2.5. This is because electricity prices and groundwater depths are seasonally correlated—groundwater levels are lowest (making pumping more expensive) in the winter months, when electricity prices are also low. This likely explains why $\hat{\beta}$ estimates in Table 2.6 are smaller than $\hat{\beta}^{\text{e}}$ and $\hat{\beta}^{\text{w}}$ estimates in Table 2.5.

2.6 Open-Access Externality

We use the above elasticity estimates to quantify the pumping cost externality for each farm in our sample. This calculates the deadweight loss created when farm i fails to internalize how its own groundwater extraction increases its neighbors' costs of extraction. Importantly, farm i only creates deadweight loss if its own surplus gained from pumping a marginal acre-foot is *less* than the sum of surplus lost by *all* neighboring farms due to marginally lower groundwater levels. We focus exclusively on the (more static) "pumping cost" externality, where marginal changes in farm i's extraction induce marginal changes neighboring farms' groundwater levels and pumping costs. We do not investigate the (more dynamic) "stock externality," where failing to internalize the full continuation value of the groundwater stock causes farmers to extract too quickly.

Figure 2.3 illustrates our framework for calculating the pumping cost externality, for the simplified case where four farms share a single aquifer. Each farm has a unique groundwater demand curve, which governs the amount of groundwater it extracts (Q_i) at a given marginal pumping cost (P_i). If Farm 1 extracts less groundwater, it incurs a loss in consumer surplus (i.e., the red triangle). However, this has the effect of increasing groundwater levels for Farms 2, 3, and 4—which leads to farm-specific decreases in their marginal pumping costs (since the translation from depth to \$/AF depends on each farm's unique electricity price and pumping technology).[27] At lower pumping costs, Farms 2, 3, and 4 increase extraction and capture additional consumer surplus (i.e., the green trapezoids). If the sum of surplus gained by Farms 2, 3, and 4 is greater than the surplus lost by Farm 1, then Farm 1's open-access extraction is creating deadweight loss.

We require two inputs to operationalize Figure 2.3. First, we must parameterize i-specific groundwater demand curves. To do this, we assume constant elasticity of demand, with a

[26] Again, it is not surprising that Equations (2.3) and (2.8) yield similar point estimates, since Q^{water} and P^{water} are multiplicative transformations of Q^{elec} and P^{elec} and both specificaitons are log-log.

[27] The increase in groundwater levels may also vary across farms sharing an aquifer, as hydrogeological factors govern the speed at which groundwater flows both horizontally and vertically through the substrata.

homogeneous elasticity of $\epsilon = -1.12$ (based on our preferred estimate from Table 2.6). Each demand curve is simply $Q = A_i P^{\epsilon}$, where A_i is calibrated to unit i's baseline groundwater price and quantity in month t. For this exercise, we focus on June–July of 2016, the two months with the greatest groundwater extraction in our last complete year of data.

Second, we must characterize how a marginal decrease in farm i's groundwater extraction affects groundwater levels at each neighboring farm j. The requires assumptions for both the spatial patterns of how groundwater moves horizontally through the substrata (i.e. at what distance from farm i do groundwater levels respond to i's extraction) and the speed at which groundwater levels reequilibrate (i.e. how quickly does farm i's extraction impact farm j's groundwater levels). We make two crude assumptions for the sake of tractability: (i) when farm i extracts 1 acre-foot less, this causes groundwater levels to increase equally at all points within a radius of r miles, where the increase is equal to 1 foot divided by the area (in acres) of the r-mile circle;[28] (ii) when farm i extracts 1 acre-foot less in June, groundwater levels throughout the r-mile circle fully reequilibrate by July. We acknowledge that these assumptions are overly simplistic, and future versions of this analysis will incorporate more realistic assumptions based on detailed hydrogeological models of California's groundwater basins (Sunding, Zilberman, and Brozovic 2010).

Table 2.7 reports deadweight loss calculations for each farm (i.e. service point) in the three most common groundwater basins in our sample: the Sacramento Valley (641 APEP-matched service points actively pumping in June 2016), the Salinas Valley (941), and the San Joaquín Valley (5,325). In each basin, the median farm would have lost about \$0.50 in consumer surplus from pumping 1 AF less in June 2016 (i.e. the red triangle from Figure 2.3). This consumer surplus loss varies substantially, as the 1st-quartile farm would have lost up to \$1.96, while the 3rd-quartile farm would have lost as little as \$0.16.[29] The bottom three panels calculate the total change in consumer surplus, including both farm i's lost surplus and the surplus gained by all neighboring farms j within an r-mile radius. We present two calculations for each groundwater basin and radius: (i) "APEP" columns calculate $\sum_j \Delta CS_j$ for all neighboring service points in our APEP-matched estimation sample; (ii) "Scaled" columns inflate the number of neighbors by the proportion of *unmatched* agricultural service points.[30] Bolded numbers indicate positive welfare changes from reducing deadweight loss—where farm i's marginal open-access externality is greater than its own private marginal surplus.

Four notable patterns emerge from the distributions of changes in deadweight loss. First, greater spatial density of groundwater pumping increases the magnitude of the open-access

[28] Aquifers are not bathtubs, and groundwater levels can vary substantially within a given r-mile circle. For this simple exercise, we assume that groundwater levels increase in parallel throughout the circle. This preserves baseline heterogeneity in groundwater levels across space.

[29] This variation in ΔCS_i reflects baseline differences in P_i and Q_i, which we use to parameterize each i-specific demand curve.

[30] Our econometric analysis includes only service points that we can directly match to an APEP pump test. However, this sample represents a relatively small subset of PGE agricultural customers—many of whom are also groundwater pumpers (groundwater pumping is the leading agricultural energy end-use in California). We calculate conservative scaling factors based on the proportion of unmatched PGE services points with the most common tariffs in the matched sample, which errs towards excluding unmatched pumpers (rather than erring towards including unmatched non-pumpers).

externality, as $\sum_j \Delta CS_j$ sums across a greater number of neighbors. This is not surprising, yet Table 2.7 highlights stark differences across basins and between "APEP" vs. "Scaled" calculations. The static pumping externality is largest in the San Joaquín Valley, which contains the majority of all PGE agricultural customers. Accounting for unmatched groundwater pumpers is also important, as the APEP matched sample dramatically undercounts the number of neighbors impacted by farm i's extraction.

Second, increasing the size of the radius r has an ambiguous effect on the size of the externality: larger circles include more neighbors but spread the additional 1 AF of groundwater across a greater area, yielding smaller cost reductions for each neighbor. Table 2.7 revels that larger cirlces tend to yield more positive welfare changes, especially in the San Joaquín Valley. Where pumping is the most geographically dense, the number of farm i's neighbors increases fast enough with r to offset increasingly tiny marginal changes in groundwater depth.

Third, we observe a substantial amount of variation in deadweight loss within each groundwater basin, radius, and sample. While the 3rd quartile of deadweight loss is positive for eight of nine "Scaled" calculations, the 1st quartile is always negative. This implies that over 25 percent of farms in each groundwater basin are likely contributing deadweight loss on the margin, but a different 25 percent of farms are likely *not* contributing any marginal deadweight loss. An optimal price- or quantity-based policy would need to account for variation in the marginal damages imposed by each farm i, though there is less cross-sectional variation in the marginal externality itself ($\sum_j \Delta CS_j$) than in the marginal deadweight loss ($\Delta CS_i + \sum_j \Delta CS_j$). To illustrate this cross-sectiional heterogeneity, Figure 2.4 plots kernel densities of marginal surplus changes for farms in the San Joaquín Valley (corresponding to the righthand columns of Table 2.7).[31]

Fourth, the magnitudes of our deadweight loss calculations are small relative to farmers' marginal pumping costs. For the 5,325 farmers in the San Joaquín Valley, their average marginal cost of groundwater pumping in June 2016 was \$60 per AF. Our largest estimates of the marginal deadweight loss are an order of magnitude smaller. At the same time, these 5,325 farmers extracted nearly 400,000 AF of groundwater in June 2016 alone. This suggests that due to farmers' elastic demand for groundwater, a proportionately small tax to internalize the pumping cost externality could have eliminated nearly \$1 million in deadweight loss, for just one summer month. Also, Table 2.7 likely understates the true magnitudes of deadweight loss by only considering ΔCS_j for July 2016. In reality, if farm i pumps 1 AF less in June, neighboring farms should continue to accrue marginal increases in consumer surplus after July.

These calculations are still preliminary, and we interpret them with caution. Once we incorporate more sophisticated hydrogeological assumptions, we plan to (i) replace simple radii with more spatially nuanced representations of each farm's neighbors; (ii) relax our assumption of parallel depth increases for all neighboring farms, to allow for groundwater equilibration to occur at different speeds for different neighbors; and (iii) relax our crude

[31] These kernel densities omit farms above (below) the top (bottom) 10 percent of the ΔCS_i distribution, for ease of presentation. Hence, they understate the full cross-sectional heterogeneity within the San Joaquín Valley.

treatment of timing, to allow neighbors to accrue losses in surplus for as many months as the short-run change in depth is likely to persist. We also plan to apply machine learning techniques to predict which unmatched farms are groundwater pumpers, rather than simply scaling by the rate of unmatched service points. As an input to this prediction exercise, we will assign all service points to polygons for Common Land Units (i.e. agricultural fields) and parcels (i.e. contiguous farms under common ownership), to remove the subset of unmatched services points located far from agricultural crop lands.

Assigning services points to polygons will also let us aggregate up to the more natural cross-sectional unit—the farm. The will prevent us from mistaking multiple pumps on the same farm for neighbors. We also plan to overlay the USDA's Cropland Data Layer, in order to assign a specific crop (or set of crops) to each farm. This will allow us to estimate heterogeneous demand elasticities by crop type. While our current calculations are driven entirely by differences in A_i and variation in the spatial density of i's neighbors, incorporating heterogeneous ϵ_i will allow us to capture differences between annuals vs. perennials, crops with high vs. low water intensity, and high- vs. low-value crops. For example, almond farmers (perennial, high water intensity, high value) are likely much more inelastic groundwater users than alfalfa farmers (perennial, high water intensity, low value), due to almonds' higher marginal value product per acre-foot of irrigated water. It follows that alfalfa farmers are more likely to impose large pumping externalities on almond farmers than vice versa.

2.7 Conclusion

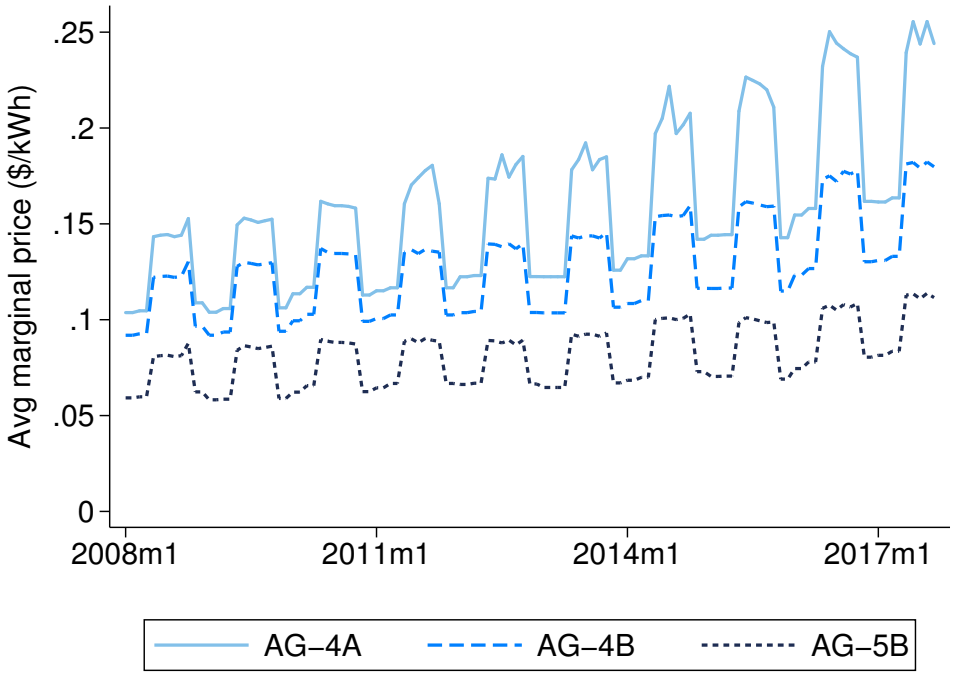
This paper estimates the price elasticities of demand both for agricultural electricity consumption (for groundwater pumping) and for agricultural groundwater consumption. We overcome common data challenges by linking customer-specific electricity consumption data with data from in-person pump efficiency audits, in order construct accurate measures for farmers' groundwater consumption and pumping costs. We credibly identify our elasticity estimates using exogenous variation in agricultural electricity tariffs, and our results reveal farmers to be much more elastic groundwater pumpers than previously thought. Using these elasticity estimates, we quantify the "pumping cost" externality—or the extent to which farm i increases its neighbors' groundwater extraction costs by removing a marginal unit of water from their shared aquifer. Our preliminary results suggest that the magnitude of this externality is likely small relative to the private costs of groundwater pumping, and that a small corrective tax could yield substantial welfare gains for California farmers.

These results are still preliminary, and future versions of this paper will incorporate more sophisticated hydrogeological assumptions to more accurately characterize the spatial nature of the pumping cost externality. We also plan to use detailed geospatial data in order to aggregate up from electricity service points to farms, as well as high-resolution satellite imagery to classify farms by crop type. This will allow us to estimate heterogeneous elasticities by crop type, and to incorporate these heterogeneous elasticities into our externality calculations. A richer version of Table 2.7 will differentiate between annuals vs. perennials, crops with high vs. low water intensity, and high- vs. low-value crops—letting us characterize the extent to which the pumping cost externality contributes to misallocation in crop choice.

Future work will use quasi-experimental variation in farmers' pumping capital to estimate the long-run elasticity of demand for groundwater.

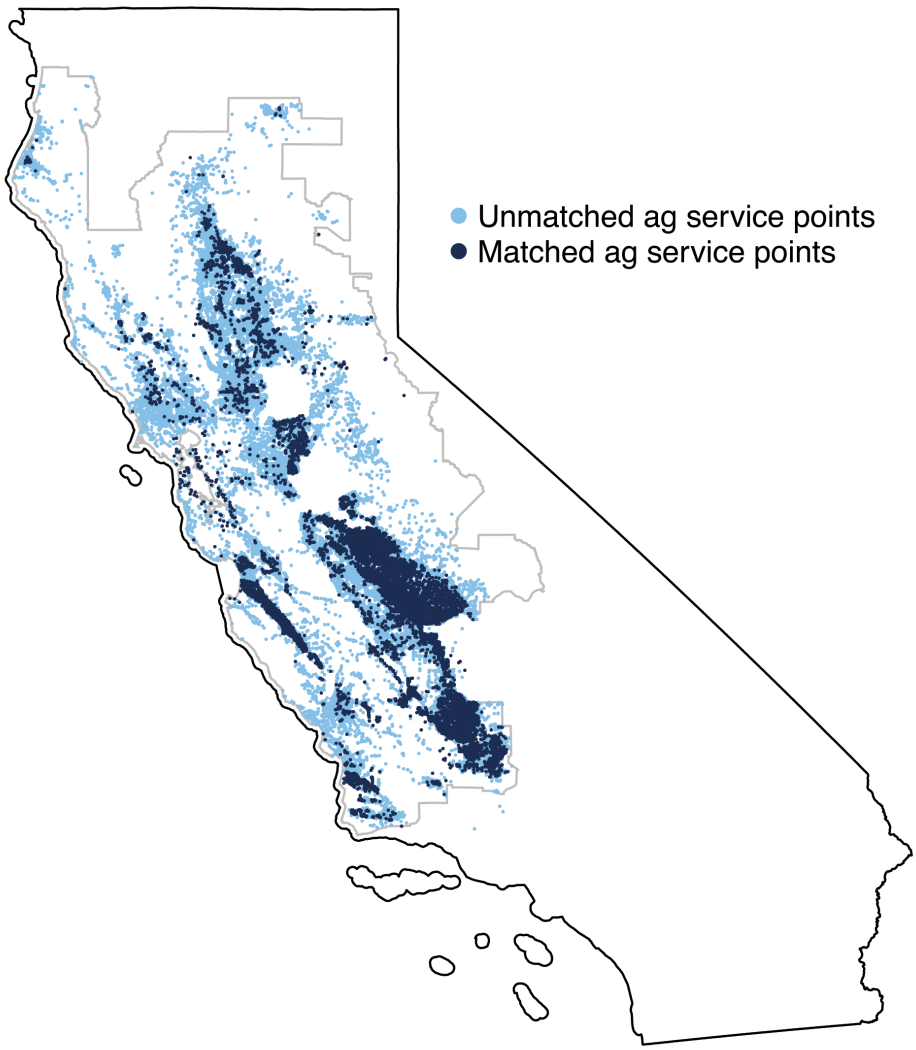
Figures

Figure 2.1: Average Marginal Electricity Prices



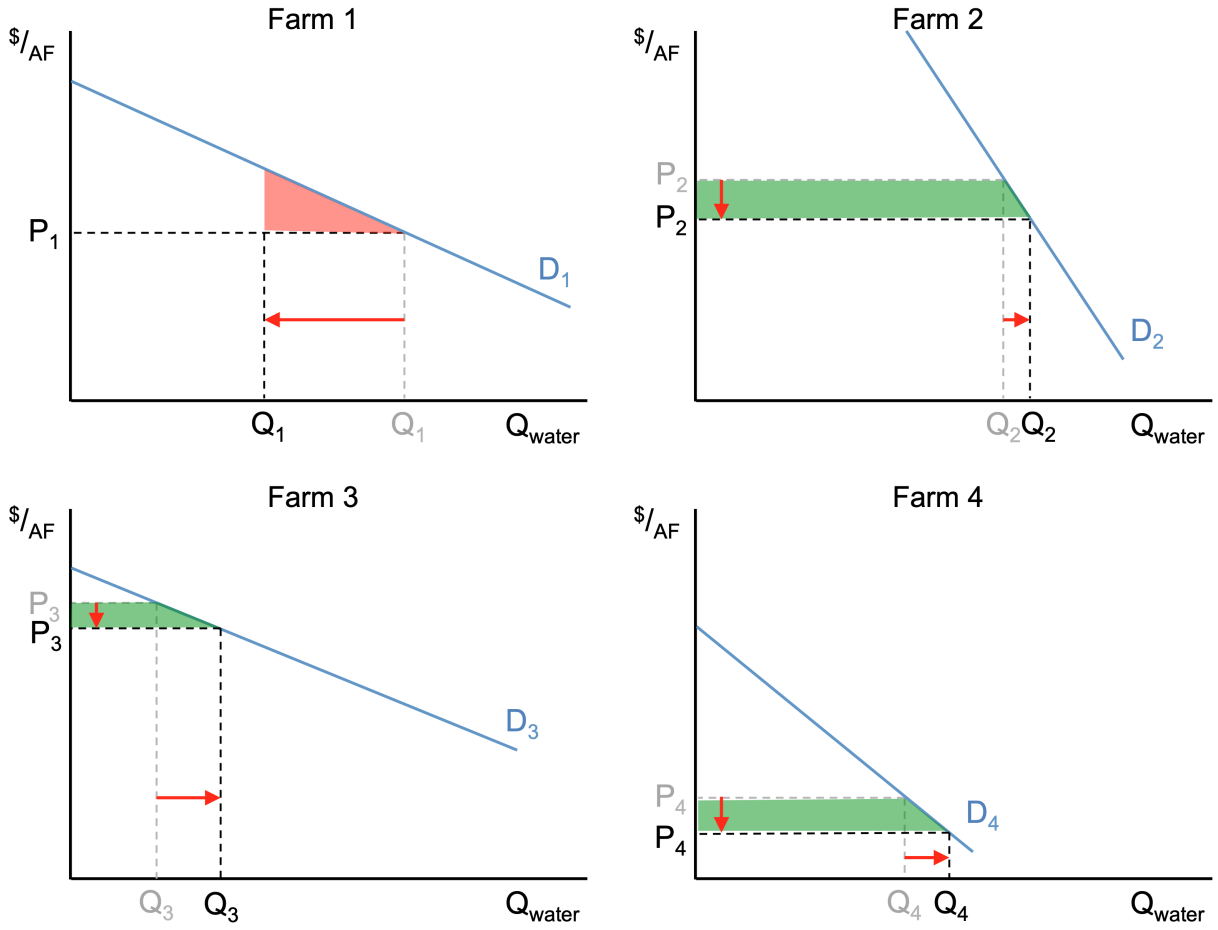
Notes: This figure reports the times series of monthly average marginal electricity prices ($/kWh) for three of the most common agricultural tariffs. Prices are systematically higher during summer months (May–October). Much of our identifying variation in monthly electricity prices comes from these monthly price times series not rising perfectly in parallel.

Figure 2.2: PGE Agricultural Customers



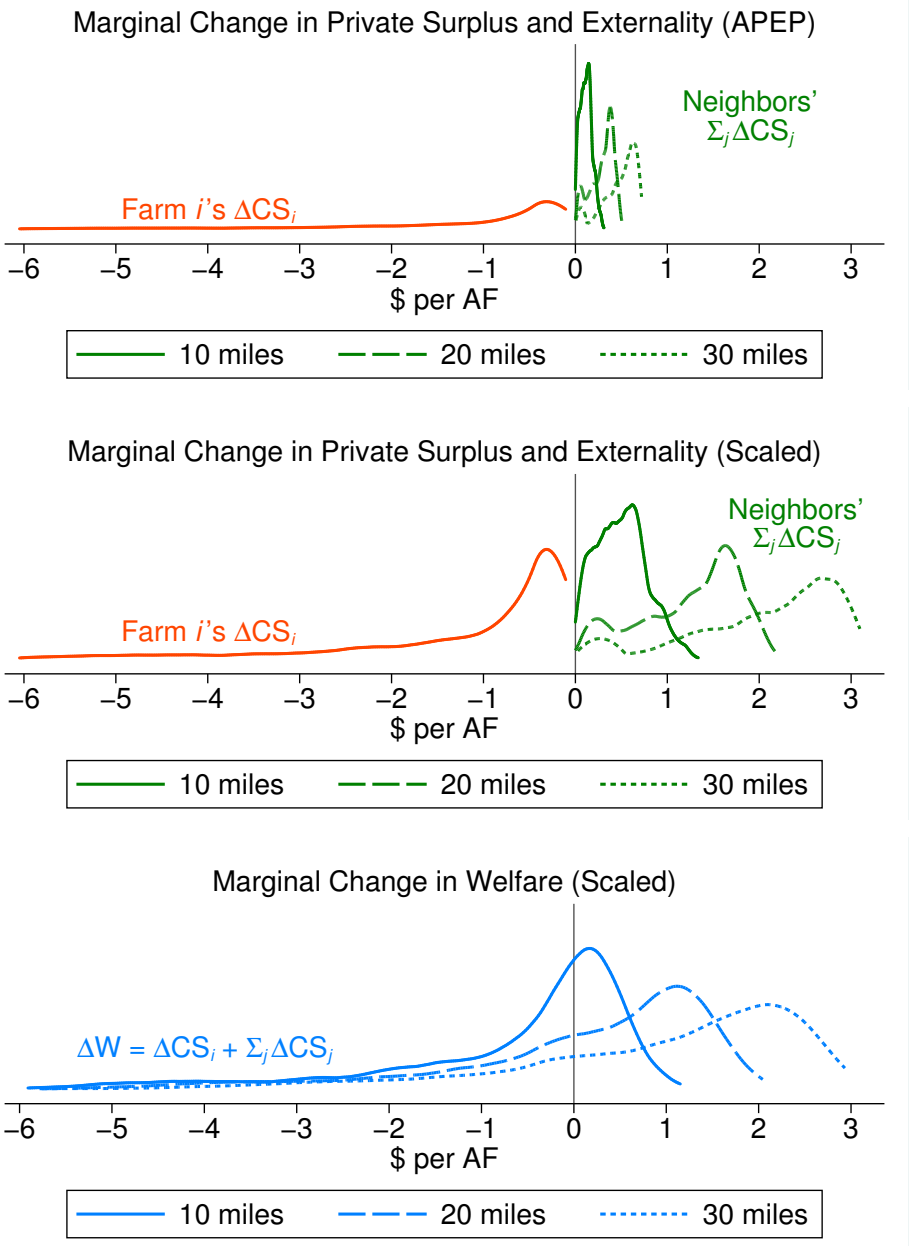
Notes: This figure maps the locations of all agricultural service points served by PGE. Dark blue dots indicate the 11,851 service point that we can match directly to an APEP pump test. Light blue dots indicate unmatched agricultural service points. The light grey outline is the geographic boundary of PGE's service territory.

Figure 2.3: Framework for Calculating the Pumping Cost Externality



Notes: This figure illustrates how we conceptualize and calculate the pumping cost externality that farm i imposes on its neighbors—by not internalizing how its own groundwater extraction increases extraction costs for other nearby farms. In this example, Farm 1 has three neighbors (Farms 2, 3, and 4), and all four farms have distinct demand curves for groundwater. If Farm 1 pumps less than its desired quantity of groundwater, its consumer surplus falls (illustrated by the red triangle). However, when Farm 1 extracts less groundwater, this slightly increases groundwater levels for neighboring Farms 2, 3, and 4. Because Farms 2, 3, and 4 have distinct pumping technologies and electricity prices, a parallel increase in groundwater levels leads to idiosyncratic decreases in their effective prices of groundwater. Farms 2, 3, and 4 respond to these price decreases by slightly increasing groundwater extraction, leading to increased consumer surplus (illustrated by the green trapezoids). If the sum of the (green) surplus gains for all of Farm 1's neighbors (i.e. the pumping cost externality) is greater than Farm 1's own (red) surplus loss, then the Social Planner can increase welfare by taxing Farm 1's groundwater extraction.

Figure 2.4: Pumping Cost Externality: Densities for San Joaquín Valley



Notes: This figure plots kernel densities for marginal changes in surplus, for APEP farms in the San Joaquín Valley (as reported in the righthand columns of Table 2.7). Red densities show the distribution of farm i's change in private surplus from pumping 1 AF less in June 2016. Green densities show distributions of surplus gains for farm i's neighbors (within r-mile circles) in July 2016, which are not reported in Table 2.7. Blue densities sum the orange and green densities to show the net change in welfare (as in the bottom 3 panels of Table 2.7); positive mass indicates deadweight loss, where farm i's marginal pumping cost externality is larger than its marginal private surplus. The middle and bottom densities scale up our APEP-matched sample to include unmatched agricultural electricity consumers likely to be groundwater pumpers. All densities omit farms outside the 10th-to-90th percentiles of ΔCS_i, for ease or presentation. See text for further details.

Tables

Table 2.1: PGE Agricultural Tariffs

Category	Tariff	Description	Percent
Small pumps, conventional meters single motor < 35 hp, or multiple motors summing to < 15 hp	**1A**	High price per kWh (not time-varying), fixed charge per hp connected	3.0
Large pumps, conventional meters single motor ≥ 35 hp, or multiple motors summing to ≥ 15 hp, or single overloaded motor ≥ 15 hp	**1B**	High price per kWh (not time-varying), fixed charge per max kW consumed	8.1
Small pumps, smart meters single motor < 35 hp, or multiple motors summing to < 15 hp	**4A** (4D)	High prices per kWh (higher in peak hours), fixed charges per hp connected, very high peak prices on 14 summer Event Days	7.2
	5A (5D)	Lower prices per kWh (peak & offpeak), no Event Day price increases, higher fixed charges per hp	2.7
	RA (RD)	Lower peak prices per kWh, higher off-peak prices per kWh, no Event Day price increases, choice between MTW or WTF peak days	1.2
	VA (VD)	Lower peak prices per kWh, higher off-peak prices per kWh, no Event Day price increases, choice of 3 shorter 4-hour peak periods	0.9
Large pumps, smart meters single motor ≥ 35 hp, or multiple motors summing to ≥ 15 hp, or single overloaded motor ≥ 15 hp	**4B** (4E)	High prices per kWh (higher in peak hours), fixed charges per max kW consumed	20.1
	5B (5E)	Much lower prices per kWh (peak & offpeak), higher fixed charge per max kW	37.8
	4C (4F)	Slightly lower prices per kWh (peak & offpeak), higher fixed charges per kW shifted to peak, very high peak prices on 14 summer Event Days	2.4
	5C (5F)	Much lower prices per kWh (peak & offpeak), higher fixed charges per kW shifted to peak, very high peak prices on 14 summer Event Days	7.8
	RB (RE)	Higher prices per kWh (peak & off-peak), choice between MTW or WTF peak days, lower fixed charges per max kW (in summer)	1.5
	VB (VF)	Higher prices per kWh (peak & off-peak), choice of 3 shorter 4-hour peak periods, lower fixed charges per max kW (in summer)	0.6
Customers transitioning off internal combustion engines	**ICE**	Very low price per kWh (high in peak hours), fixed charge per max kW consumed	6.8

Notes: This table provides a rough summary of PGE's 23 electricity tariffs for agricultural customers. The first column lists the 5 disjoint categories of customers, defined (primarily) by physical pumping capital and electricity meters. Effective default tarrifs within each group are in bold, and farmers may switch tariffs *within* a category (but not *across* categories). All tariffs have fixed ($/kW) and volumetric ($/kWh) prices that vary by summer vs. winter. All time-of-use tariffs (i.e. all but 1A and 1B) also vary between peak (12:00pm–6:00pm on summer weekdays), partial peak (8:30am–9:30pm on weekends), and off-peak periods. DEF tariffs are functionally equivalent to their ABC analogs, and are holdovers for the earliest customers to adopt time-of-use pricing. Actual tariffs are far more complex, and tariff documents are available at `https://www.pge.com/tariffs/index.page`. The right-most column reports the percent of observations in our main estimation sample on each tariff.

Table 2.2: Summary Statistics – Electricity Data

	All Ag Customers	Matched to Pumps
Service point-month observations	9,991,458	1,168,553
Unique service points (SPs)	108,172	11,851
SPs that switch tariff categories	44,414	2,844
SPs that switch categories (pumping capital)	3,454	561
SPs that switch categories (smart meters)	43,045	2,553
Share of SP-months on time-varying tariffs	0.702	0.886
Share of SP-months on peak-day tariffs	0.295	0.152
Monthly electricity consumption (kWh)	6080.9	12055.7
	(39783.1)	(25075.1)
Monthly electricity consumption (kWh), summer	8249.6	17589.1
	(45660.8)	(29818.5)
Monthly electricity consumption (kWh), winter	3849.7	6362.8
	(32498.8)	(17232.5)
Average marginal electricity price ($/kWh)	0.148	0.113
	(0.050)	(0.042)
Average marginal electricity price ($/kWh), summer	0.171	0.130
	(0.051)	(0.044)
Average marginal electricity price ($/kWh), winter	0.126	0.096
	(0.037)	(0.032)
Average monthly bill ($, non-zero bills)	936.66	1814.15
	(4662.71)	(3285.26)
Average monthly bill ($, non-zero bills), summer	1398.90	2821.16
	(5847.34)	(4020.99)
Average monthly bill ($, non-zero bills), winter	456.17	764.99
	(2888.86)	(1742.21)

Notes: The left column reports summary statistics for the universe of agricultural electricity customers in PGE service territory, from 2008–2017. The right column includes the subset of agricultural customers that we successfully match to a groundwater pump in the APEP pump test dataset—i.e., our main estimation sample. "Pumping capital" denotes tariff category switches driven by shifts between small pumps (< 35 hp) and large pumps (≥ 35 hp), or adding/removing an auxiliary internal combustion engine. Most tariff category switches were driven by PGE's smart meter rollout. Time-varying tariffs (i.e. all except 1A and 1B) have higher marginal prices during peak demand hours. Peak-day tariffs (i.e. 4A, 4D, 4C, 4F, 5C, 5F) have very high marginal prices during peak hours on the 14 highest-demand summer days. Monthly bills include both volumetric ($/kWh) and fixed charges ($/kW, $/hp, and $/day). Summer months are May–October. Standard deviations of sample means in parentheses.

Table 2.3: Summary Statistics – Pump Tests and Groundwater Consumption

	Matched to Pumps
Service point-month observations	1,168,553
Unique service points (SPs)	11,851
Matched APEP points per SP	3.45
	(8.80)
Operating pump efficiency (%)	54.46
	(11.52)
kWh per AF conversion factor (APEP measured)	430.30
	(254.21)
kWh per AF conversion factor (constructed)	346.94
	(206.03)
Monthly groundwater consumption (AF)	49.0
	(151.7)
Monthly groundwater consumption (AF), summer	74.0
	(191.9)
Monthly groundwater consumption (AF), winter	23.4
	(86.6)
Average marginal groundwater price ($/AF)	39.91
	(29.63)
Average marginal groundwater price ($/AF), summer	41.76
	(31.34)
Average marginal groundwater price ($/AF), winter	38.01
	(27.63)

Notes: These summary stats are from the merged panel of groundwater prices and quantities, which combines electricity data, pump test data, and groundwater data. We observe 3.45 unique APEP pump tests for the average matched service point, although 37 percent of service points match to only a single APEP test. Our constructed kWh per AF conversion factor (i.e. $\widehat{\text{kWh/AF}}_{it}$) uses monthly groundwater rasters to capture changes in (measured) kWh per AF over time, and estimation error compresses the right tail of distribution of measured kWh per AF. Monthly groundwater consumption divides electricity consumption (kWh) by $\widehat{\text{kWh/AF}}_{it}$. Grounwater prices multiply marignal electricity prices ($/kWh) by $\widehat{\text{kWh/AF}}_{it}$. Summer months are May–October. Standard deviations of sample means in parentheses.

Table 2.4: Estimated Demand Elasticities – Electricity

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
$\log(P_{it}^{\text{elec}})$	-1.31^{***}	-1.58^{***}	-1.17^{***}	-1.02^{***}	-1.18^{***}	-0.76^{***}
	(0.11)	(0.17)	(0.16)	(0.14)	(0.21)	(0.17)
Instrument(s):						
Default $\log(P_{it}^{\text{elec}})$		Yes	Yes	Yes		Yes
Default $\log(P_{it}^{\text{elec}})$, lagged					Yes	
Fixed effects:						
Unit × month-of-year	Yes	Yes	Yes	Yes	Yes	Yes
Month-of-sample	Yes	Yes	Yes	Yes	Yes	Yes
Unit × physical capital			Yes	Yes	Yes	Yes
Water basin × year				Yes		
Water district × year				Yes		
Unit-specific linear time trends						Yes
Service point units	11,175	11,175	11,175	11,121	10,924	11,175
Months	117	117	117	117	105	117
Observations	1.05M	1.05M	1.05M	1.04M	0.91M	1.05M
First stage F-statistic		4136	7382	7508	757	4776

Notes: Each regression estimates Equation (2.3) at the service point by month level, where the dependent variable is the inverse hyperbolic sine transformation of electricity consumed by service point i in month t. We estimate IV specifications via two-stage least squares, instrumenting with either unit i's within-category default logged electricity price in month t *or* the 6- and 12- month lags of this variable. "Physical capital" is a categorical variable for (i) small pumps, (ii) large pumps, and (iii) internal combustion engines, and unit × physical capital fixed effects control for shifts in tariff category triggered by the installation of new pumping equipment. Water basin × year fixed effects control for broad geographic trends in groundwater depth. Water district × year fixed effects control for annual variation in surface water allocations. All regressions drop solar NEM customers, customers with bad geocodes, and months with irregular electricity bills (e.g. first/last bills, bills longer/shorter than 1 month, overlapping bills for a single account). Standard errors (in parentheses) are clustered by service point and by month-of-sample. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2.5: Estimated Demand Elasticities – Groundwater

	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
$\log(P_{it}^{\text{elec}})$: $\hat{\beta}^{\text{e}}$	-1.21^{***}	-1.39^{***}	-1.38^{***}	-1.26^{***}	-1.23^{***}	-1.68^{***}
	(0.17)	(0.18)	(0.19)	(0.17)	(0.16)	(0.21)
$\log\left(\widehat{\frac{\text{kWh}}{\text{AF}}}_{it}\right)$: $\hat{\beta}^{\text{w}}$	-0.92^{***}	-1.37^{***}	-1.32^{***}	-1.72^{***}	-1.24^{***}	-2.04^{***}
	(0.11)	(0.25)	(0.28)	(0.30)	(0.24)	(0.45)
Instrument(s):						
Default $\log(P_{it}^{\text{elec}})$	Yes	Yes	Yes	Yes	Yes	Yes
log (Avg depth in basin)		Yes	Yes	Yes	Yes	
log (Avg depth in basin), lagged						Yes
Fixed effects:						
Unit × month-of-year	Yes	Yes	Yes	Yes	Yes	Yes
Month-of-sample	Yes	Yes	Yes	Yes	Yes	Yes
Unit × physical capital	Yes	Yes	Yes	Yes	Yes	Yes
Water basin × year					Yes	
Water district × year					Yes	
Groundwater time step	Month	Month	Month	Quarter	Month	Month
Only basins with > 1000 SPs			Yes			
Service point units	10,159	10,121	9,324	10,134	10,086	9,890
Months	117	116	116	117	116	105
Observations	0.93M	0.83M	0.80M	0.89M	0.83M	0.70M
First stage F-statistic	6932	129	144	61	69	20

Notes: Each regression estimates Equation (2.7) at the service point by month level, where the dependent variable is the inverse hyperbolic sine transformation of electricity consumed by service point i in month t. We report estimates for $\hat{\beta}^{\text{e}}$ and $\hat{\beta}^{\text{w}}$, where the latter subtracts 1 from the estimated coefficient on $\log\left(\widehat{\text{kWh}/\text{AF}}_{it}\right)$. We estimate IV specifications via two-stage least squares, and all regressions instrument for P_{it}^{elec} with unit i's within-category default logged electricity price in month t (consistent with our preferred specification from Table 2.4). We instrument for $\log\left(\widehat{\text{kWh}/\text{AF}}_{it}\right)$ with either logged average groundwater depth across unit i's basin, or the 6- and 12-month lags of this variable. "Physical capital" is a categorical variable for (i) small pumps, (ii) large pumps, and (iii) internal combustion engines, and unit × physical capital fixed effects control for shifts in tariff category triggered by the installation of new pumping equipment. Water basin × year fixed effects control for broad geographic trends in groundwater depth. Water district × year fixed effects control for annual variation in surface water allocations. Column (3) restricts the sample to only the three most common water basins (San Joaquin Valley, Sacramento Valley, and Salinas Valley), each of which contains over 1000 unique SPs in our estimation sample. Column (4) uses a quarterly panel of groundwater depths to construct $\log\left(\widehat{\text{kWh}/\text{AF}}_{it}\right)$ and the instrument, rather than a monthly panel. All regressions drop solar NEM customers, customers with bad geocodes, months with irregular electricity bills (e.g. first/last bills, bills longer/shorter than 1 month, overlapping bills for a single account), and pumps with implausible test measurements. Standard errors (in parentheses) are clustered by service point and by month-of-sample. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2.6: Estimated Demand Elasticities – Groundwater

	(1) OLS	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
$\log(P_{it}^{\text{water}})$	-0.81^{***}	-1.12^{***}	-1.16^{***}	-1.12^{***}	-0.97^{***}	-1.14^{***}
	(0.09)	(0.15)	(0.17)	(0.15)	(0.14)	(0.21)
Instrument(s):						
Default $\log(P_{it}^{\text{elec}})$		Yes	Yes	Yes	Yes	
Default $\log(P_{it}^{\text{elec}})$, lagged						Yes
Fixed effects:						
Unit × month-of-year	Yes	Yes	Yes	Yes	Yes	Yes
Month-of-sample	Yes	Yes	Yes	Yes	Yes	Yes
Unit × physical capital	Yes	Yes	Yes	Yes	Yes	Yes
Water basin × year					Yes	
Water district × year					Yes	
Groundwater time step	Month	Month	Month	Quarter	Month	Month
Only basins with > 1000 SPs			Yes			
Service point units	10,159	10,159	9,342	10,159	10,118	9,926
Months	117	117	117	117	117	105
Observations	0.93M	0.93M	0.85M	0.93M	0.93M	0.82M
First stage F-statistic		2835	2735	2846	4633	486

Notes: Each regression estimates Equation (2.8) at the service point by month level, where the dependent variable is the inverse hyperbolic sine transformation of groudnwater consumed by service point i in month t. We estimate IV specifications via two-stage least squares, and Columns (2)–(5) instrument for P_{it}^{water} with unit i's within-category default logged electricity price. Column (6) instruments with the 6- and 12- month lags of this variable. "Physical capital" is a categorical variable for (i) small pumps, (ii) large pumps, and (iii) internal combustion engines, and unit × physical capital fixed effects control for shifts in tariff category triggered by the installation of new pumping equipment. Water basin × year fixed effects control for broad geographic trends in groundwater depth. Water district × year fixed effects control for annual variation in surface water allocations. Column (3) restricts the sample to only the three most common water basins (San Joaquin Valley, Sacramento Valley, and Salinas Valley), each of which contains over 1000 unique SPs in our estimation sample. Column (4) uses a quarterly panel of groundwater depths to construct both Q_{it}^{water} and P_{it}^{water}, rather than a monthly panel. All regressions drop solar NEM customers, customers with bad geocodes, months with irregular electricity bills (e.g. first/last bills, bills longer/shorter than 1 month, overlapping bills for a single account), and pumps with implausible test measurements. Standard errors (in parentheses) are clustered by service point and by month-of-sample. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2.7: Deadweight Loss Calculations for June 2016

	Sacramento Valley		Salinas Valley		San Joaquín Valley	
APEP units (i)	641		964		5,325	
ΔCS_i from 1 AF less						
25th percentile	−1.05		−1.61		−1.96	
50th percentile	−0.45		−0.59		−0.56	
75th percentile	−0.16		−0.28		−0.22	
10-mile radius	APEP	Scaled	APEP	Scaled	APEP	Scaled
Mean # of neighbors (j)	112	1,207	257	590	363	1,563
$\Delta CS_i + \sum_j \Delta CS_j$						
25th percentile	−1.03	−0.86	−1.54	−1.46	−1.86	−1.52
50th percentile	−0.43	−0.24	−0.50	−0.39	−0.45	−0.12
75th percentile	−0.14	**0.00**	−0.18	−0.05	−0.10	**0.29**
20-mile radius	APEP	Scaled	APEP	Scaled	APEP	Scaled
Mean # of neighbors (j)	236	2,554	498	1,145	1,047	4,500
$\Delta CS_i + \sum_j \Delta CS_j$						
25th percentile	−1.02	−0.66	−1.51	−1.32	−1.67	−0.90
50th percentile	−0.42	−0.09	−0.45	−0.28	−0.28	**0.51**
75th percentile	−0.13	**0.12**	−0.12	**0.08**	**0.07**	**1.14**
30-mile radius	APEP	Scaled	APEP	Scaled	APEP	Scaled
Mean # of neighbors (j)	339	3,659	710	1,632	1,736	7,463
$\Delta CS_i + \sum_j \Delta CS_j$						
25th percentile	−1.01	−0.54	−1.45	−1.23	−1.47	−0.21
50th percentile	−0.41	**0.05**	−0.41	−0.18	−0.11	**1.32**
75th percentile	−0.12	**0.27**	−0.07	**0.21**	**0.27**	**2.06**

Notes: This table reports calculations for the open-access externality, for service point i in our main estimation sample located in the three largest groundwater basins (Sacramento Valley, Salinas Valley, and San Joaquín Valley). This simple exercise includes only service points with positive groundwater extraction in June 2016, with counts reported in the top row. For each unit i, we calculate their private decrease in consumer surplus from pumping 1 AF less in June, 2016 (reported in the top panel). Next, we translate unit i's 1-AF lower extraction in June into the corresponding marginal increase in groundwater levels in July across i-centered circles of radius r. For neighboring units j within each circle, this marginal increase in groundwater levels translates to a marginal decrease in j's effective price of groundwater. The bottom three panels report the net effect on *total* consumer surplus, subtracting i's lost consumer surplus in June from increased consumer surplus in July summed across all units j. Bolded numbers indicate positive welfare changes, consistent with unit i imposing a negative open-access externality greater than its own private benefit. . "APEP" columns include only neighbors in our APEP-matched estimation sample, which almost certainly understates the magnitude of $\sum_j \Delta CS_j$ (by summing over only a subset of nearby agricultural groundwater pumpers). "Scaled" columns conservately inflate the number of i's neighbors based on the ratio of match-to-unmatched PGE agricultural service points in each groundwater basin. We calculate all changes in consumer surplus by parameterizing unit-specific groundwater demand curves, imposing a homogeneous and constant elasticity of $\epsilon = -1.12$ (based on our estimate from Table 2.6, Column (2)). All units are in \$/AF.

Chapter 3

Panel Data and Experimental Design[1]

3.1 Introduction

Randomized controlled trials (RCTs) are an increasingly popular method for applied economics research (Card, DellaVigna, and Malmendier 2011). When designing RCTs, researchers typically use *ex ante* power calculations to evaluate the trade-off between sample size and statistical precision. If the sample is too small, the experiment will be unable to distinguish between true and false null hypotheses; at the same time, overly large samples waste resources.[2] The economics literature on power calculations has focused on single-wave experiments, where units are randomized into treatment and control groups, and researchers observe each unit once.[3] In a widely cited paper based on results from Frison and Pocock (1992), McKenzie (2012) recommends panel data experiments, where multiple observations per unit help to increase statistical power. This is especially attractive in settings where collecting additional waves of data for one individual is cheaper than enrolling more individuals. As a result, panel RCTs have become increasingly common in recent years.[4]

[1]This chapter is coauthored with Fiona Burlig and Louis Preonas. The original version can be found online at `https://ei.haas.berkeley.edu/research/papers/WP%20277.pdf`. We thank Andrew Foster, several anonymous refereees, Michael Anderson, Maximilian Auffhammer, Patrick Baylis, Joshua Blonz, Severin Borenstein, Hai-Anh Dang, Aureo de Paula, Meredith Fowlie, James Gillan, Erin Kelley, Jeremy Magruder, Aprajit Mahajan, Edward Miguel, Tavneet Suri, Catherine Wolfram, and seminar participants at the University of California Berkeley, BITSS, and NEUDC for valuable comments. Molly Van Dop provided excellent research assistance.

[2]Bloom (1995) provides an early framework for power calculations, while Duflo, Glennerster, and Kremer (2007) and Glennerster and Takavarasha (2013) detail their implementation in practice. Cohen (1977) and Murphy, Myors, and Wolach (2014) are also classic references.

[3]Researchers often collect two waves of data, but estimate treatment effects using post-treatment data only, controlling for the baseline level of the outcome variable (following McKenzie (2012)). Baird et al. (2018) extends the standard cross-sectional setup to randomized saturation designs, capable of measuring spillover and general equilibrium effects. Athey and Imbens (2017b) discusses statistical power using a randomization inference approach.

[4]Recent panel RCTs include Bloom et al. (2013), Blattman, Fiala, and Martinez (2014), Jessoe and Rapson (2014), Bloom et al. (2015), Fowlie et al. (2017), Atkin, Khandelwal, and Osman (2017), Atkin et al. (2017), McKenzie (2017), and Fowlie, Greenstone, and Wolfram (2018).

At the same time, panel data pose challenges for statistical inference, due to non-constant within-unit serial correlation. Bertrand, Duflo, and Mullainathan (2004) demonstrate that failing to account for this correlation structure can bias standard errors towards zero, raising the probability of a Type I error. In order to achieve correct false rejection rates, applied econometricians often implement the cluster-robust variance estimator (CRVE), or use "clustered standard errors", which accommodates arbitrary serial correlation within panel units.[5]

Besides affecting the false rejection rate, serial correlation can also impact statistical power. Hence, it is important to properly account for serial correlation during both *ex post* analysis and *ex ante* power calculations.[6] However, McKenzie (2012) only allows for *constant* serial correlation *ex ante* (equivalent to assuming i.i.d. errors after removing unit fixed effects), despite evidence that panel data typically exhibit more complex serial correlation (Bertrand, Duflo, and Mullainathan 2004). While Frison and Pocock (1992) propose a framework for incorporating non-constant serial correlation into *ex ante* power calculations, they rely on restrictive assumptions—homogeneous error structures across units and deterministic time shocks—which are likely unrealistic in practice. As a result, there is a gap in the existing literature that may preclude researchers from designing properly powered experiments using panel data.

In this paper, we derive analytical expressions for the variance of panel estimators under *arbitrary* non-i.i.d. error structures, while also allowing for random time shocks. We use these expressions to formalize a power calculation formula for difference-in-differences estimators that is robust to arbitrary serial correlation.[7]

We conduct Monte Carlo analyses using both simulated and real data, and demonstrate that the methods outlined in McKenzie (2012) yield experiments that are incorrectly powered in the presence of non-constant serial correlation, even with proper *ex post* inference. These methods tend to yield dramatically overpowered experiments in short panels and dramatically underpowered experiments in long panels. By allowing for non-constant serial correlation *ex ante*, our "serial-correlation-robust" power calculation approach achieves the desired power in both simulated and real data. Ultimately, we provide researchers with both the theoretical insights and practical tools to design properly powered experiments in panel data settings.

[5]See White (1984), Arellano (1987), and Cameron and Miller (2015) for more details on the CRVE. Abadie et al. (2017) underscore the need to use the CRVE in experiments where treatment assignment is correlated within clusters—as in most panel RCTs, where treated units remain treated throughout the experiment.

[6]If researchers do not adjust their standard errors *ex post* to account for within-unit serial correlation in panel data, they will likely over-reject true null hypotheses. If they adjust their standard errors *ex post* but do not adjust their *ex ante* power calculations to account for within-unit serial correlation, they introduce a mismatch between *ex ante* and *ex post* assumptions that will likely yield incorrectly powered experiments.

[7]Recent experiments published in top economics journals use either the difference-in-differences estimator or the ANCOVA estimator. We discuss ANCOVA in Sections 3.2.2.2 and 3.3.1, where standard power calculation techniques similarly ignore non-constant serial correlation within panel units. However, analytically deriving the variance of the ANCOVA estimator necessitates restrictive assumptions on the data generating process, causing analytical ANCOVA power calculations to perform poorly with real data. For this reason, we focus on power calculations for difference-in-differences.

We make three main contributions to the economics literature on experimental design. First, we show that standard power calculation methods for panel RCTs (discussed in McKenzie (2012)) fail in the presence of arbitrary serial correlation. Second, we derive a new power calculation formula for difference-in-differences, allowing for *arbitrary* serial correlation, which extends Frison and Pocock (1992) by accommodating heterogeneous error structures across units and random time shocks. This serial-correlation-robust formula enables researchers to calibrate panel RCTs to the desired power. Finally, we provide guidance for designing panel RCTs in real experimental settings, and introduce an accompanying STATA package (`pcpanel`) to facilitate proper power calculations in practice.

The paper proceeds as follows. Section 3.2 presents analytical power calculations for panel data with arbitrary serial correlation, and uses Monte Carlo simulations to evaluate the performance of these results. Section 3.3 extends these simulation results to real experimental data. Section 3.4 discusses practical issues related to power calculations, and introduces our accompanying software package. Section 3.5 concludes.

3.2 Power calculations for panel data

Power calculations provide an *ex ante* estimate of the smallest effect size that an experiment, with a given sample size and experimental design will be able to statistically detect. Most power calculations take the following form:

$$MDE = \left(t^d_{1-\kappa} + t^d_{\alpha/2}\right) \sqrt{\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right)} \tag{3.1}$$

where $\text{Var}(\hat{\tau} \mid \mathbf{X})$ is the exact finite sample variance of the treatment effect estimator, conditional on independent variables $\mathbf{X}$; $t^d_{\alpha/2}$ is the critical value of a t distribution with d degrees of freedom associated with the probability of a Type I error, α, in a two-sided test against a null hypothesis of $\tau = 0$; and $t^d_{1-\kappa}$ is the critical value associated with the probability of correctly rejecting a false null, κ.[8] These parameters determine the minimum detectable effect (MDE), the smallest value $|\tau| > 0$ for which the experiment will (correctly) reject the null $\tau = 0$ with probability κ at the significance level α.

This paper's core contribution is our "serial-correlation-robust" (SCR) power calculation formula for designing experiments with panel data. This section outlines our model and resulting SCR formula for the difference-in-differences (DD) estimator, which extends Frison and Pocock (1992) and McKenzie (2012) by incorporating arbitrary (non-constant) serial correlation. Using simulations, we demonstrate the importance of accounting for this correlation *ex ante* in order to achieve the desired statistical power *ex post*. We also consider three sensitivities: short panels, which are traditionally of interest in development economics; alternative treatment effect estimators; and alternative assumptions on the data generating process (DGP).

[8] For one-sided tests, $t^d_{\alpha/2}$ can be replaced with t^d_{α}. $1 - \kappa$ gives the probability of a false rejection, or a Type II error. The degrees of freedom, d, will depend on the dimensions of $\mathbf{X}$ and the treatment effect estimator in question.

3.2.1 Serial-correlation-robust power calculations

We begin with a model in which there are J units, P proportion of which are randomized into treatment. The researcher collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.[9]

Assumption 1 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

where υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$; and τ is the treatment effect, which is assumed to be homogeneous across all units and all time periods.[10]

Assumption 2 (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J - 1$ *unit dummies, and* $(m + r) - 1$ *time dummies. This follows from random assignment of* D_{it}*.*

Assumption 3 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption 4 (Independence across units)**.** $\mathbb{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0, \ \forall \ i \neq j, \ \forall \ t, s.$

Assumption 5 (Symmetric covariance structures)**.** *Define:*

$$\psi^B \equiv \frac{2}{Jm(m-1)} \sum_{i=1}^{J} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^A \equiv \frac{2}{Jr(r-1)} \sum_{i=1}^{J} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^X \equiv \frac{1}{Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of the same unit, respectively. Define ψ_T^B, ψ_T^A, and ψ_T^X analogously, where we

[9] Put differently, we assume that there is a control group of units that is never treated in the sample period, and a treatment group of units for which treatment turns on in a particular time period (and persists through all subsequent periods). This is a standard design for panel RCTs in economics.

[10] While we assume a homogeneous treatment effect, our derivations also hold if τ varies across time periods (as in Frison and Pocock (1992) and McKenzie (2012)). Section 3.2.2.3 explores data generating processes where υ_i and δ_t are deterministic, rather than random shocks.

consider only the PJ treated units; also define ψ_C^B, ψ_C^A, and ψ_C^X analogously, where we consider only the $(1-P)J$ control units. Using these definitions, assume that $\psi^B = \psi_T^B = \psi_C^B$; $\psi^A = \psi_T^A = \psi_C^A$; and $\psi^X = \psi_T^X = \psi_C^X$.[11]

Under these assumptions, the OLS estimator with unit and time fixed effects is $\widehat{\tau} = (\ddot{\mathbf{D}}'\ddot{\mathbf{D}})^{-1}\ddot{\mathbf{D}}'\ddot{\mathbf{Y}}$, with $\mathbb{E}[\hat{\tau}] = \tau$. Assumptions 1–5 yield a power calculation formula that is robust to arbitrary serial correlation:[12]

$$MDE = (t_{1-\kappa}^J + t_{\alpha/2}^J)\sqrt{\underbrace{\left(\tfrac{1}{P(1-P)J}\right)\left[\left(\tfrac{m+r}{mr}\right)\sigma_\omega^2 + \left(\tfrac{m-1}{m}\right)\psi^B + \left(\tfrac{r-1}{r}\right)\psi^A - 2\psi^X\right]}_{\text{Var}(\hat{\tau}|\mathbf{X})}} \tag{3.2}$$

Throughout the remainder of the paper, we refer to Equation (3.2) as the "serial-correlation-robust" (SCR) power calculation formula. Note that under cross-sectional randomization, this expression for the variance of $\hat{\tau}$ still holds in expectation, even in the presence of within-period error correlations across units:

Lemma 1. *In a panel difference-in-differences model with treatment randomly assigned at the unit level, $\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_\omega^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]$ is an unbiased estimator of the expectation of $\text{Var}(\hat{\tau} \mid \mathbf{X})$, even in the presence of arbitrary within-period cross-sectional correlations. See Appendix B.1.3 for a proof, and see Appendix B.1.2.3 for a more general model that relaxes Assumptions 4–5.*

3.2.1.1 Accounting for serial correlation

Our SCR power calculation formula generalizes the Frison and Pocock (1992, henceforth FP) difference-in-differences formula to accomodate fully arbitrary correlation structures. Whereas FP assume that all units share a homogeneous correlation structure, our SCR formula accommodates arbitrary heterogeneous correlation structures across units. We also allow for random time shocks across all units, which are characteristic of most economic datasets.[13] As a result, our SCR formula is able to flexibly accommodate more realistic panel data structures.

McKenzie (2012) is the most widely cited reference for power calculations using panel data in economics. McKenzie's results follow from FP's original derivations, yet they impose a more restrictive assumption on the data generating process: constant serial correla-

[11] These ψ terms come from the derivation of the variance of the DD estimator (see Appendix B.1.2.2). Note that they depend both on the error structure (ω_{it} from the underlying DGP) and on the length of the experiment (m and r). We choose the letters "B" to indicate the Before-treatment period, and "A" to indicate the After-treatment period. We index the m pre-treatment periods $\{-m+1, \ldots, 0\}$, and the r post-treatment periods $\{1, \ldots, r\}$. In a randomized setting, $\mathbb{E}\left[\psi^B\right] = \mathbb{E}\left[\psi_T^B\right] = \mathbb{E}\left[\psi_C^B\right]$, $\mathbb{E}\left[\psi^A\right] = \mathbb{E}\left[\psi_T^A\right] = \mathbb{E}\left[\psi_C^A\right]$, and $\mathbb{E}\left[\psi^X\right] = \mathbb{E}\left[\psi_T^X\right] = \mathbb{E}\left[\psi_C^X\right]$, making this a reasonable assumption *ex ante*. However, it is possible for treatment to alter the covariance structure of treated units only.

[12] We present the formal derivation of this formula in Appendix B.1.2.2. Note that if $m = 1$ (or $r = 1$), ψ^B (or ψ^A) is not defined and is multiplied by 0 in Equation (3.2).

[13] If we impose FP's assumptions, our SCR formula collapses to the FP formula for the difference-in-differences model with unequal correlations.

tion between any two time periods, within each cross-sectional unit.[14] Using our notation, McKenzie's difference-in-differences power calculation formula becomes:[15]

$$MDE = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\frac{\sigma^2_\omega}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)} \tag{3.3}$$

This is equivalent to the SCR formula when ψ^A, ψ^B, and ψ^X are all equal to zero—allowing only *constant* serial correlation of the *composite* error term (ε_{it}, in McKenzie's notation), or an i.i.d. idiosyncratic error term (ω_{it}, in our notation). As highlighted by Bertrand, Duflo, and Mullainathan (2004, henceforth BDM), this assumption is likely unrealistic because most panel datasets in economics exhibit non-constant serial correlation.[16]

To further illustrate the difference between the McKenzie and SCR models, consider two cross-sectional units (indexed $\{i, j\}$) and four time periods (indexed $\{0, 1, 2, 3\}$), with the data generating process described in Assumption 1. The vector of idiosyncratic errors, $\boldsymbol{\omega}$, and the corresponding variance-covariance matrix, $\boldsymbol{\Omega}$, can be represented as follows:[17]

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_{i0} \\ \omega_{i1} \\ \omega_{i2} \\ \omega_{i3} \\ \omega_{j0} \\ \omega_{j1} \\ \omega_{j2} \\ \omega_{j3} \end{bmatrix} \qquad \boldsymbol{\Omega} = \begin{bmatrix} \sigma^2_{i0} & & & & & & & \\ \sigma_{i0,i1} & \sigma^2_{i1} & & & & & & \\ \sigma_{i0,i2} & \sigma_{i1,i2} & \sigma^2_{i2} & & & & & \\ \sigma_{i0,i3} & \sigma_{i1,i3} & \sigma_{i2,i3} & \sigma^2_{i3} & & & & \\ & & & & \sigma^2_{j0} & & & \\ & & & & \sigma_{j0,j1} & \sigma^2_{j1} & & \\ & & & & \sigma_{j0,j2} & \sigma_{j1,j2} & \sigma^2_{j2} & \\ & & & & \sigma_{j0,j3} & \sigma_{j1,j3} & \sigma_{j2,j3} & \sigma^2_{j3} \end{bmatrix}$$

Serial correlation within each unit is represented by the (potentially non-zero) covariance terms $\sigma_{it,is}$ and $\sigma_{jt,js}$, for all $t \neq s$. In contrast, McKenzie's model allows only for constant serial correlation—that is, $\upsilon_i \neq 0$ and ω_{it} i.i.d., in our notation. The assumption of i.i.d. idiosyncratic errors would force all off-diagonal covariance elements in $\boldsymbol{\Omega}$ to equal zero, thereby precluding the types of non-constant serial correlation that BDM highlight (e.g., autoregressive processes).

The magnitudes of these off-diagonal covariance terms directly affect the variance of the DD estimator. The three ψ terms defined above, along with the error variance and experimental design parameters, are sufficient to fully characterize the true variance of the

[14] The preceding paragraph describes the most general FP model, reported on page 1701. McKenzie draws from the more restrictive FP model, reported on page 1693.

[15] See Appendix B.1.2.1 for the derivation.

[16] Our SCR formula differs from that of McKenzie in one additional way: McKenzie's empirical specification does not include a unit fixed effect, whereas the estimator underlying the SCR model does. This fixed effect absorbs any *constant* serial correlation of composite errors, leaving no remaining serial correlation between the idiosyncratic error terms of our SCR model. Thus, constant serial correlation in McKenzie's framework translates to *no* serial correlation in our SCR framework. Section 3.2.2.3 discusses the implications of removing *constant* serial correlation from the true DGP, and Section 3.2.2.2 considers alternative treatment effect estimators.

[17] We show only the lower diagonal of the variance-covariance matrix because $\boldsymbol{\Omega}$ is symmetric. We also omit the cross-unit covariance terms for notational convenience, which are zero under Assumption 4.

treatment effect estimator in this model. To fix ideas, using the four-period model above and supposing treatment is administered beginning at $t = 2$, these covariance parameters are:

$$\psi^B = \frac{\sigma_{i0,i1} + \sigma_{j0,j1}}{2}$$
$$\psi^A = \frac{\sigma_{i2,i3} + \sigma_{j2,j3}}{2}$$
$$\psi^X = \frac{\sigma_{i0,i2} + \sigma_{i1,i2} + \sigma_{i0,i3} + \sigma_{i1,i3} + \sigma_{j0,j2} + \sigma_{j1,j2} + \sigma_{j0,j3} + \sigma_{j1,j3}}{8}$$

Alternatively, if treatment is administered beginning at $t = 1$, these covariance terms become:

$$\psi^B = \text{(not defined for only 1 pre-treatment period)}$$
$$\psi^A = \frac{\sigma_{i1,i2} + \sigma_{i1,i3} + \sigma_{i2,i3} + \sigma_{j1,j2} + \sigma_{j1,j3} + \sigma_{j2,j3}}{6}$$
$$\psi^X = \frac{\sigma_{i0,i1} + \sigma_{i0,i2} + \sigma_{i0,i3} + \sigma_{j0,j1} + \sigma_{j0,j2} + \sigma_{j0,j3}}{6}$$

Assumption 5 generalizes this structure to a model with J units across m pre-treatment periods and r post-treatment periods. Equation (3.2) shows that greater average covariance in the pre- or post-treatment periods (ψ^B or ψ^A) increases the MDE. Intuitively, as errors for treated and control units are more serially correlated, the benefits of collecting multiple waves of pre- and post-treatment data are eroded. However, cross-period covariance (ψ^X) enters Equation (3.2) negatively. This highlights a key property of the DD estimator: because DD identifies the treatment effect off of differences between post- and pre-treatment outcomes, greater serial correlation between pre- and post-treatment observations makes differences caused by treatment easier to detect.

Assuming that the within-unit correlation structure does not vary systematically across time periods, positively correlated errors will imply positive ψ^B, ψ^A, and ψ^X. Because ψ^B and ψ^A enter the SCR formula positively, while ψ^X enters negatively, serial correlation may either increase or decrease the MDE relative to the McKenzie i.i.d. case. Specifically, serial correlation will increase the MDE if and only if:

$$\left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A > 2\psi^X \tag{3.4}$$

This inequality is more likely to hold in longer panels, for two reasons. First, as the number of pre- and post-treatment periods increases, $\left(\frac{m-1}{m}\right)$ and $\left(\frac{r-1}{r}\right)$ approach one. Second, the covariance terms contributing to ψ^X lie farther away from the diagonal of the variance-covariance matrix than the covariance terms contributing to ψ^B and ψ^A. Because errors from non-adjacent time periods are likely to be less correlated than errors from adjacent time periods, and because the number of far-off-diagonal covariances increases relatively more quickly for ψ^X as the panel becomes longer, ψ^X is increasingly likely to be smaller than ψ^B and ψ^A in longer panels.[18]

[18]This analytical result illustrates how the correlation structure impacts the variance of estimators that use both pre- and post-treatment data. Figure 3.10 below demonstrates that, in cases with strong serial correlation, adding time periods can actually increase the MDE. This is not specific to the DD estimator: Appendix Figure B.4 shows the same pattern for power calculations using the ANCOVA estimator.

3.2.1.2 Monte Carlo simulations

If a randomized experiment relies on a power calculation that fails to properly account for serial correlation *ex ante*, its realized power may be different from the desired κ. To understand the extent to which this matters in practice, we conduct a series of Monte Carlo simulations comparing the McKenzie model and the SCR model over a range of panel lengths and error correlations. We simulate three cases and compute the Type I error rate and the statistical power for each: (i) experiments that fail to account for serial correlation in ω_{it} (i.e. the *idiosyncratic* component of error term) both *ex ante* and *ex post*; (ii) experiments that fail to account for serial correlation in ω_{it} *ex ante* but apply the CRVE to account for serial correlation *ex post*; and (iii) experiments that both account for serial correlation in ω_{it} *ex ante* and apply the CRVE *ex post*.

For each set of parameter values characterizing both a data generating process and an experimental design, we first calculate two treatment effect sizes: τ^{McK} equal to the MDE from the McKenzie formula, and τ^{SCR} equal to the MDE from our SCR formula. Second, we use these parameter values to create a panel dataset from the following DGP:

$$Y_{it} = \beta + \upsilon_i + \delta_t + \omega_{it} \tag{3.5}$$

where ω_{it} follows an AR(1) process:

$$\omega_{it} = \gamma\omega_{i(t-1)} + \xi_{it} \tag{3.6}$$

Third, we randomly assign treatment, with effect sizes τ^{McK}, τ^{SCR}, and $\tau^0 = 0$ at the unit level, to create three separate outcome variables. Fourth, we regress each of these outcome variables on their respective treatment indicators and include unit fixed effects and time fixed effects. Fifth, we compute both OLS standard errors and CRVE standard errors clustered at the unit level, for all three regressions. We repeat steps two through five 10,000 times for each set of parameters, calculating rejection rates of the null hypothesis $\tau = 0$ across all simulations. For τ^{McK} and τ^{SCR}, this rate represents the realized power of the experiment. For the placebo τ^0, it represents the realized false rejection rate.

We test five levels of the AR(1) parameter: $\gamma \in \{0, 0.3, 0.5, 0.7, 0.9\}$. For each γ, we simulate symmetric panels with an equal number of pre-treatment and post-treatment periods, with panel lengths ranging from 2 periods ($m = r = 1$) to 40 periods ($m = r = 20$). We hold J, P, β, σ^2_υ, σ^2_δ, α, and κ fixed across all simulations, and we adjust the variance of the white noise term σ^2_ξ such that every simulation has a fixed idiosyncratic variance σ^2_ω. This allows γ to govern the proportion of σ^2_ω that is serially correlated.[19] The covariance terms ψ^B, ψ^A, and ψ^X have closed-form expressions under the AR(1) structure, and we use these expressions to calculate τ^{SCR}.[20] This causes τ^{SCR} to vary both with the degree of serial correlation and panel length, whereas τ^{McK} varies only with panel length.

Figure 3.1 displays the results of this exercise. The left column shows rejection rates under the McKenzie formula using OLS standard errors, which assumes zero serial correlation in

[19] In an AR(1) model, the relationship between the variance of the AR(1) process and the variance of the white noise disturbance depends on γ, with $\sigma^2_\omega = \frac{\sigma^2_\xi}{1-\gamma^2}$.

[20] Appendix B.2.1 provides these derivations, along with further details on these Monte Carlo simulations.

ω_{it} both *ex ante* and *ex post*. The middle column shows rejection rates under the FP formula using CRVE standard errors, which accounts for serial correlation in ω_{it} *ex post* only. The right column show rejection rates under our SCR formula using CRVE standard errors, which allows for serial correlation in ω_{it} both *ex ante* and *ex post*. The top row plots realized power as a function of the number of pre/post-treatment periods, which should equal $\kappa = 0.80$ in a properly designed experiment. The bottom row plots the corresponding realized false rejection rates, which should equal the desired $\alpha = 0.05$. Only the SCR formula, in conjunction with CRVE standard errors, achieves the desired 0.80 and 0.05 across all panel lengths and AR(1) parameters.

The left column confirms the BDM result that failing to appropriately account for serial correlation leads to false rejection rates dramatically higher than $\alpha = 0.05$. Even a modest AR(1) parameter of $\gamma = 0.5$ yields a 20 percent probability of a Type I error, for panels with $m = r > 5$. This underscores the fact that randomization cannot correct serial correlation in panel settings, and experiments that collect multiple waves of data from the same cross-sectional units should account for within-unit correlations over time. By contrast, the middle and right columns apply the CRVE and reject placebo effects at the desired rate of $\alpha = 0.05$.

The middle column shows how failing to properly account for serial correlation *ex ante* can yield dramatically overpowered or underpowered experiments. Particularly for longer panels with $m = r > 5$, performing power calculations via Equation (3.3) may actually produce experiments with less than 50 percent power, even though researchers intended to achieve power of 80 percent (i.e., $\kappa = 0.80$). For a relatively high serial correlation of $\gamma = 0.7$, simulations based on the conventional power calculation formula yield power less than 32 percent for $m = r > 10$. This is consistent with the BDM finding that applying the CRVE reduces statistical power, even though doing so achieves the desired Type I error rate. By contrast, the right column applies both the SCR power calculation formula and the CRVE, and these simulations achieve the desired power of $\kappa = 0.80$ for each value of γ.

The middle column also highlights how failing to account for serial correlation *ex ante* may either increase *or* decrease statistical power, as shown in Equation (3.4). For shorter panels, using the FP formula instead of our SCR formula yields dramatically overpowered experiments. While this may seem counterintuitive, Equation (3.4) is increasingly unlikely to hold as m and r decrease to 1. In the extreme case where $m = r = 1$, ψ^B and ψ^A do not enter, and the only covariance term in the SCR formula is ψ^X, which enters negatively. These simulations reveal that just as higher γ yields more dramatically underpowered experiments for longer panels, higher γ yields more dramatically *over*powered experiments for shorter panels.[21]

These results are striking. For even a modest degree of AR(1) serial correlation, applying the McKenzie power calculation formula will not yield experiments of the desired statistical power. By contrast, the SCR formula achieves the desired power for all panel lengths and AR(1) parameters. While AR(1) is a relatively simple correlation structure, it serves as a reasonable first-approximation for more complex forms of serial correlation. Given that real-

[21]Intuitively, serial correlation has two opposite effects on the statistical power of a DD estimator. It decreases power by reducing the effective number of observations for each cross-sectional unit, and it increases power by increasing the signal in estimating treatment effects off of a post−pre difference. In shorter panels, this second effect tends to dominate.

world panel datasets exhibit enough serial correlation to produce high Type I error rates, it stands to reason that such serial correlation can similarly impact the statistical power of experiments if not properly accounted for *ex ante*.

3.2.2 Sensitivities

3.2.2.1 Short panels

While experiments with long panels have become increasingly common, short-panel experiments remain a development economics staple. Our SCR formula generalizes to panels as short as 2 periods ($m = r = 1$), where Equation (3.2) simplifies to:

$$MDE = (t^J_{1-\kappa} + t^J_{\alpha/2})\sqrt{\left(\frac{1}{P(1-P)J}\right)[2\sigma^2_\omega - 2\psi^X]} \tag{3.7}$$

Notably, ψ^B and ψ^A are no longer defined for $m = r = 1$. However, ψ^X remains, meaning that the wedge between the SCR and McKenzie formulas also remains. By omitting ψ^X (which enters Equation (3.7) negatively), the McKenzie formula will tend to yield over-powered two-period experiments.

It may be unintuitive that serial correlation still matters in a two-period panel, which is isomorphic to a cross-sectional first-differences model, and does not require clustered standard errors to achieve the desired Type I error rate. However, unlike the Type I error rate, statistical power (i.e. 1 minus the Type II error rate) is a function of the variance of the treatment effect estimator, which depends on ψ^X. Hence, ignoring ψ^X from Equation (3.7) will bias MDE upward, likely leading researchers to choose an unnecessarily large J.[22]

To illustrate the difference between the SCR and McKenzie formulas for short panel experiments, we extend Figure 3.1 for panels with between 1 and 6 pre/post-treatment periods. We again simulate data with serially correlated errors of varying AR(1) parameters γ, and calibrate treatment effect sizes using each power calculation formula. Figure 3.2 displays realized power for panels with $1 \leq m \leq 6$ and $1 \leq r \leq 6$, varying m and r independently. Across all levels of non-zero serial correlation ($\gamma > 0$), the McKenzie formula yields over-powered experiments for panels with either 1 pre-treatment or 1 post-treatment period. Consistent with Figure 3.1, realized power under the McKenzie formula decreases monotonically as panel length increases; the longer the panel, the more likely the McKenzie formula is to deliver underpowered experiments. While the McKenzie formula yields overpowered short panel experiments using these simulated data, it can yield *underpowered* experiments in real data, even for short panels.[23] By contrast, experiments calibrated using the SCR formula achieve the desired power regardless of panel length.

3.2.2.2 Alternative estimators

Our SCR power calculation formula assumes that researchers will employ a DD estimator with unit and time fixed effects. However, particularly in experiments with short panels,

[22] Appendix B.1.2.4 mathematically demonstrates how power depends on ψ^X even in a two-period panel, despite the fact that the false rejection rate in a two-period panel is independent of ψ^X.

[23] See Appendix Figure B.2.

researchers often choose to use one of two alternative approaches: a simplified differences-in-differences estimator or an analysis of covariance (ANCOVA) estimator.

Simplified differences-in-differences Researchers often estimate DD models that replace a full set of unit and time fixed effects with a treatment group dummy and/or post-period dummy, respectively. The resulting estimating equation takes the following form:

$$Y_{it} = \beta + \tau \left[\text{Treat}_i \times \text{Post}_t\right] + \upsilon \left[\text{Treat}_i\right] + \delta \left[\text{Post}_t\right] + \varepsilon_{it} \tag{3.8}$$

where $\text{Treat}_i = 1$ if unit i is in the treatment group, and $\text{Post}_t = 1$ for all post-treatment periods. Figure 3.3 explores how the McKenzie and SCR formulas perform under three alternative *ex post* DD estimating equations, which alter our original DD estimating equation by: (i) replacing unit fixed effects with a Treat_i dummy, (ii) replacing time fixed effects with a Post_t dummy, and (iii) replacing both sets of fixed effects with dummies (Equation (3.8)). In all cases, we use the same DGP as in Assumption 1, which includes idiosyncratic unit and time shocks; as before, we simulate data following Equation (3.5), with an idiosyncratic error term following an AR(1) process.

As in Figure 3.1, the SCR formula achieves correctly powered experiments at all levels of serial correlation, for all three DD estimators in Figure 3.3. This is because all three yield the same treatment effect estimator—and therefore have the same variance—as our original DD estimating equation with unit and time fixed effects. However, these different estimating equations have varying levels of serial correlation in the *error term*. Using OLS standard errors would produce different estimates of the variance, even though the true underlying variance is the same across these estimating equations. On the other hand, with randomized treatment, the CRVE (clustered at the unit level) correctly estimates the (same) variance in all cases, because it accounts for serial correlation within unit, regardless of whether that correlation enters through the underlying error structure or the omission of fixed effects.[24] Realized power under the McKenzie formula is likewise equivalent across all three alternative estimators, yielding overpowered experiments in short panels and underpowered experiments in long panels.

ANCOVA Another common strategy, particularly in short panels, is to employ the analysis of covariance (ANCOVA) estimator.[25] To do this, the econometrician estimates the following specification using post-treatment data only:

$$Y_{it} = \alpha + \tau D_i + \theta \overline{Y}_i^B + \varepsilon_{it} \tag{3.9}$$

[24] Appendix B.1.4 mathematically shows that the DD estimator in Equation (3.8) has the same variance as the DD estimator with unit and time fixed effects.

[25] In a randomized setting where unit fixed effects are not needed for identification, this method may be preferred to DD because it more efficiently estimates $\hat{\tau}$ (Frison and Pocock 1992). McKenzie (2012) notes that under constant serial correlation (i.i.d. ω_{it} in our SCR framework), ANCOVA is always more efficient than the DD model with the same number of time periods, but that these gains are eroded as the intracluster correlation coefficient increases. These gains are also eroded as the number of pre-treatment periods increases. Neither Frison and Pocock (1992) nor McKenzie (2012) handles the fully general case of arbitrary serial correlation. Teerenstra et al. (2012) begins with a similar setup for the ANCOVA framework, but considers the $m = r = 1$ case only, obviating the need to address the CRVE-related issues raised here.

where $\overline{Y}_i^B = \sum_{t=-m+1}^{0} Y_{it}$ is the pre-treatment average value of the dependent variable for unit i. This estimator has become popular in economics, as it is more efficient than the DD model with the same number of periods (McKenzie 2012).

Frison and Pocock (1992) also derive a power calculation formula for the ANCOVA estimator, based on the same assumptions as their DD formula. As with the DD estimator, McKenzie draws on a more restrictive FP formula, which assumes constant serial correlation within each panel unit. McKenzie's formula (henceforth McKenzie ANCOVA) has become the standard for ANCOVA power calculations, which we adapt to our notation:

$$MDE \approx (t^J_{1-k} + t^J_{\alpha/2})\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2\right]} \tag{3.10}$$

where $\theta = \frac{m\sigma_v^2}{m\sigma_v^2+\sigma_\omega^2}$.[26]

Importantly, deriving this formula under non-constant within-unit serial correlation necessitates an additional simplifying assumption for analytical tractability: we must assume away time shocks.[27] Under this assumption, we derive the variance of the ANCOVA estimator allowing for arbitrary serial correlation, along with the corresponding serial-correlation-robust power calculation formula (henceforth SCR ANCOVA):

$$\begin{aligned} MDE \approx{}& (t^J_{1-\kappa} + t^J_{\alpha/2})\times \\ & \sqrt{\frac{1}{P(1-P)J}\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]} \end{aligned} \tag{3.11}$$

where $\theta = \frac{m\sigma_v^2+m\psi^X}{m\sigma_v^2+\sigma_\omega^2+(m-1)\psi^B}$.[28]

[26] Our McKenzie ANCOVA formula differs slightly from that in McKenzie (2012) and Frison and Pocock (1992, page 1693). These previous derivations have assumed that the true data generating process follows Equation (3.9), where post-treatment outcomes are determined in part by pre-treatment outcomes. We instead assume unit-specific random effects, as in Assumption 1. FP and McKenzie assumed deterministic time shocks with no variance, or $\sigma_\delta^2 = 0$. We instead must assume that there are no time shocks for analytical tractability. While both data generating processes yield identical treatment effect estimators, they imply different variances of this estimator. As in Frison and Pocock (1992) and McKenzie (2012), our McKenzie ANCOVA formula is approximate because we ignore sampling error in the estimation of θ, which approaches zero as the number of units increases.

[27] Frison and Pocock (1992) and McKenzie (2012) both assume away random time shocks. A critical step in the derivation of the ANCOVA model with time shocks and arbitrary serial correlation requires us to calculate a conditional expectation that depends on the error term ε_{it} and the pre-period mean $\overline{Y}_i^B$ of *every* unit in the experiment, which becomes analytically intractable for any reasonable number of experimental units. See Appendix B.1.2.5 for more details. By contrast, the variance of the DD estimator depends on the distribution of errors conditional on only the treatment indicator, which is orthogonal to the error terms by randomization.

[28] We present the formal derivation of the SCR ANCOVA formula in Appendix B.1.2.5. For analytical tractability, we assume that the ψ parameters are uniform across all units. We also ignore sampling error in the estimation of θ, which approaches zero as the number of units increases; Frison and Pocock (1992) and McKenzie (2012) also make this simplification. Through additional Monte Carlo simulations, we confirm that neither of these assumptions is likely to affect statistical power.

Figure 3 compares the McKenzie vs. SCR ANCOVA formulas, using simulations analogous to Figure 3.1. For each level of the AR(1) parameter and panel length, we compute the treatment effect size τ implied by each ANCOVA formula. Unlike Figures 3.1–3.3, these simulations use a data-generating process *without* random time shocks (i.e. $\sigma^2_\delta = 0$), for consistency with the assumptions underlying Equations (3.10)–(3.11). The McKenzie ANCOVA formula produces properly-powered experiments only when idiosyncratic errors are i.i.d., while the SCR ANCOVA is robust to all levels of AR(1) serial correlation.

Even though the SCR ANCOVA formula outperforms the McKenzie ANCOVA formula in the presence of AR(1) errors, we do not recommend that researchers use this formula in real-world applications (even in short panels). Time shocks with non-zero variance are a common feature of panel data, and assuming them away may result in improperly powered experiments. We discuss this further in Section 3.3 below, where we apply our SCR methods to real data. If researchers plan to use the ANCOVA estimator, we recommend that they perform power calculations by simulation, as discussed in Section 3.4.2.2 below.

3.2.2.3 Alternative data generating processes

In addition to considering alternative estimators, we also investigate the effectiveness of the SCR power calculation formula under alternative DGPs. We consider three cases: (i) deterministic unit effects ($\sigma^2_\upsilon = 0$), (ii) deterministic time effects ($\sigma^2_\delta = 0$), and (iii), deterministic unit *and* time effects ($\sigma^2_\upsilon = \sigma^2_\delta = 0$). Given that each is simply a special case of the DGP in Assumption 1, we expect the SCR to continue to achieve the desired power. However, these cases provide more favorable conditions to test the McKenzie formula: in the absence of random unit or time shocks, the variance of the *composite* error term is the same as the variance of the *idiosyncratic* error term (ω_{it}).

We test the relative performance of the SCR and McKenzie formulas in Figure 3.5. The top row presents simulations using the three alternative DGPs, while holding the fixed effects estimator fixed. In the bottom row, we match each DGP with its "matched" estimator (e.g., for the DGP with deterministic unit effects, we replace unit fixed effects with a Treat_i dummy). The SCR formula yields correctly powered experiments, even under alternative DGPs with misspecified *ex post* estimators. In contrast, the McKenzie formula generates improperly powered experiments even in the absence of unit and time shocks. Next, we test the performance of the SCR formula when we do not control the DGP, using real-world panel data.

3.3 Applications to real-world data

3.3.1 Bloom et al. (2015) data

In this section, we conduct an analogous simulation exercise using a real dataset from an experiment in a developing-country setting. These data come from Bloom et al. (2015), in which Chinese call center employees were randomly assigned to work either from home or

from the office for a nine-month period.[29] The authors estimate the following equation to derive the central result, reported in Table 2 in the original paper:

$$\text{Performance}_{it} = \alpha \text{Treat}_i \times \text{Experiment}_t + \beta_t + \gamma_i + \varepsilon_{it} \tag{3.12}$$

This is a standard DD estimating equation with fixed effects for individual i and week t. From this model's residuals, we estimate an AR(1) parameter of $\hat{\gamma} = 0.233$, which is highly statistically significant and indicates that these worker performance data exhibit weak serial correlation.

We perform Monte Carlo simulations on this dataset that are analogous to those presented above. We subset consecutive periods of the Bloom et al. (2015) dataset to create panels ranging in length from 2 periods ($m = r = 1$) to 20 periods ($m = r = 10$). For each simulation panel length, we randomly assign three treatment effect sizes, τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR}, at the individual level and estimate Equation (3.12) separately for each treatment effect size. We calibrate τ^{McK} using the McKenzie formula that assumes no serial correlation; $\tau^{AR(1)}$ using the SCR formula with ψ parameters consistent with an AR(1) error structure of $\gamma = 0.233$; and τ^{SCR} using the SCR formula with non-parametrically estimated $\psi_{\hat{\omega}}$ parameters. We define $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$ and $\psi^X_{\hat{\omega}}$ to be the estimated analogues of σ^2_{ω}, ψ^B, ψ^A, and ψ^X, where the subscript $\hat{\omega}$ denotes the variance/covariance of *residuals* rather than errors.[30]

Figure 3.6 reports the results of this exercise, demonstrating that only the SCR formula achieves the desired statistical power in the Bloom et al. (2015) data. Failing to account for non-constant serial correlation leads to experiments that deviate dramatically from 80 percent power, even when that serial correlation is relatively weak. For an experiment with 10 pre/post-treatment periods, applying the McKenzie formula with $\kappa = 0.80$ yields an experiment with only 51 percent power. This is consistent with our results from simulated data, demonstrating that researchers can calibrate a panel RCT to 80 percent power if the *ex ante* formula properly accounts for the within-unit correlation structure of the data.[31]

ANCOVA in real data The ANCOVA estimator is more efficient than the DD estimator. As discussed in Section 3.2.2.2, for DGPs with no time shocks, our SCR ANCOVA formula can achieve properly powered experiments; however, an SCR ANCOVA formula that accommodates random time shocks is not analytically tractable. Here, we test McKenzie and SCR ANCOVA formulas using the Bloom et al. (2015) data, where the true DGP

[29] This dataset consists of weekly performance measures for the 249 workers enrolled in the experiment between January 2010 and August 2011. We keep only those individuals who have non-missing performance data for the entire pre-treatment period, leaving us with a balanced panel of 79 individuals over 48 pre-treatment weeks (a different sample from that in the paper). Our purpose with this exercise is not to comment on the statistical power of the original paper, but rather to investigate the importance of accounting for serial correlation *ex ante* in real experimental data. Appendix B.2.2 provides more information on this simulation dataset, including summary statistics.

[30] Appendix B.4.1 outlines how to estimate these parameters. For all three treatment effects, we estimate the value of $\sigma^2_{\hat{\omega}}$ estimated from residuals.

[31] Appendix B.3.1 replicates the sensitivity analyses from Section 3.2.2 using the Bloom et al. (2015) data (and the Pecan Street data discussed below). We again find that the SCR formula achieves 80 percent power for short panels and DD estimators without fixed effects, while the McKenzie formula fails in both cases.

is unknown, and may contain time shocks. Figure 3.7 displays the results of these power calculations in simulated experiments of varying length, with $m \in \{1, 5, 10\}$ and $1 \leq r \leq 10$. Neither formula consistently yields properly powered experiments. The McKenzie formula comes closest with one pre-treatment period and many post-treatment periods, but performs poorly for $m > 1$ or $r < 4$. On the other hand, the SCR ANCOVA performs reasonably well with 10 pre-treatment periods, but yields overpowered experiments with $m \leq 5$ or $r < 5$.

Given that neither formula delivers properly-powered experiments with real data, we caution against using ANCOVA power calculation formulas in practice. When researchers intend to estimate an ANCOVA model *ex post*, we suggest they conduct *ex ante* power calculations using simulation-based methods (see Section 3.4.2.2 below).[32]

3.3.2 Pecan Street data

Having demonstrated the importance of properly accounting for serial correlation using data from an RCT in a developing country setting, we now turn to a much higher-frequency dataset with higher serial correlation: household electricity consumption in the United States (Pecan Street 2016).[33] Electricity consumption data tend to exhibit high within-household autocorrelation, making them particularly well-suited for this analysis. Additionally, RCTs using energy consumption data are becoming increasingly common in economics, making our Pecan Street application relevant to this growing literature.[34]

We aggregate these data to four different temporal frequencies: hourly, daily, weekly, and monthly, each with a different correlation structures and amounts of idiosyncratic variation. This allows us to compare the McKenzie vs. SCR power calculation formula over a range of underlying error structures. We conduct Monte Carlo simulations on all four temporal frequencies, following the same procedure as the Bloom et al. (2015) simulations. Figure 3.8 shows that in all four cases, realized power sharply deviates from the desired 80 percent under both the McKenzie assumption of i.i.d. idiosyncratic errors and an assumed AR(1) structure. We achieve correctly powered experimental designs only by applying the SCR method, which accounts for the full covariance structure of the Pecan Street data.

3.3.3 Power calculations in real data

To operationalize our SCR power calculation formula in practice, researchers must assume values for σ^2_ω, ψ^B, ψ^A, and ψ^X that reflect the error structure likely to be present in (future)

[32]If researchers lack access to pre-existing data for performing simulations, one option is to apply the SCR DD power calculation formula to calibrate the sample size *ex ante*, but then estimate ANCOVA *ex post*. Since ANCOVA is (weakly) more efficient than DD, this will tend to yield overpowered experiments.

[33]Pecan Street is a research organization, based at the University of Texas at Austin, that makes high-resolution energy usage data available to academic researchers. The raw data, which are available with a login from `https://dataport.pecanstreet.org/data/interactive`, consist of hourly electricity consumption for 699 households over 26,888 hours. Appendix B.2.3 provides further detail on both these data and the ensuing simulations.

[34]For example, see Allcott (2011), Jessoe and Rapson (2014), Ito, Ida, and Tanaka (2018), Fowlie, Greenstone, and Wolfram (2018), Fowlie et al. (2017), and Allcott and Greenstone (2017). There is also a large quasi-experimental literature that uses energy consumption data.

experimental datasets. In the best case scenario, researchers have access to data that are representative of what will be collected in the field, and they can estimate these variance and covariance terms from this pre-existing dataset.[35] Plugging these estimates into the SCR formula, researchers can evaluate the tradeoffs between desired power (κ), number of units (J), pre- and post-treatment observations per unit (m, r, respectively), proportion of the population treated (P), and expected effect size (MDE).

We perform this procedure on the daily Pecan Street dataset to imitate the design of an experiment that affects household electricity consumption. We do so both assuming i.i.d. idiosyncratic errors (using the McKenzie formula) and allowing for arbitrary serial correlation (using the SCR formula). For simplicity, we consider only balanced panels of households with the same number of observations before and after treatment (i.e. $m = r$). For each panel length, we re-estimate $\sigma^2_{\hat{\omega}}$ and $\psi_{\hat{\omega}}$ terms from the daily Pecan Street data.[36]

We plot the results of this exercise in Figure 3.9. The left panel applies the McKenzie formula and assumes i.i.d. idiosyncratic errors. The right panel applies the SCR formula, using our non-parametric estimates of $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ to reflect the real error structure of the data. Each curve corresponds to experiments of a particular length (ranging from $m = r = 1$ to $m = r = 10$), and plots the number of households (J) required to achieve 80 percent power as a function of MDE. In all cases, longer panels lower J needed to achieve a given MDE. However, the McKenzie formula always calls for substantially fewer households than the SCR formula. For example, with $m = r = 10$ and an MDE of 5 percent, the McKenzie formula solves for 123 households, while the SCR formula solves for 374 households—over 3 times greater sample size. Hence, if a researcher in this setting applied the CRVE *ex post* but assumes i.i.d. idiosyncratic errors *ex ante*, he would likely include too few households to achieve the desired statistical power.

3.4 Power calculations in practice

3.4.1 Trading off units and time periods

Recruiting participants, administering treatment, and collecting data are all costly, and these implementation costs are often the limiting factor in study size. We can use the power calculation framework to conceptualize the optimal design of a panel RCT given a budget, by couching it in a simple constrained optimization problem of the following form:

$$\min_{P,J,m,r} MDE(P, J, m, r) \quad \text{s.t.} \quad C(P, J, m, r) \leq B \tag{3.13}$$

[35] Appendix B.4.1 provides details on how to estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from pre-existing data, and Appendix B.5 proves that power calculations using estimated parameters recover the same MDE in expectation as those using true parameters. The plausibility of estimating these parameters will vary across settings. Researchers with implementing partners that have access to large amounts of historical data may use these data to estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. On the other hand, this may not be possible for experiments in completely unstudied settings. See Appendix B.4.3 for more details on how to overcome a lack of pre-experimental data.

[36] We fix $P = 0.5$, $\kappa = 0.80$, and $\alpha = 0.05$. See Appendix B.2.4 for details.

where $C(P, J, m, r)$ is the cost of conducting an experiment and B is the experiment's budget.

The budget constraint creates a fundamental tradeoff between including additional units and including additional time periods in the experiment, since each comes at a cost.[37] This tradeoff also arises from differences in the marginal effects of units and time periods on the MDE. Using our SCR formula, the "elasticities" of the MDE with respect to number of units and number of time periods are:

$$\frac{\partial MDE/MDE}{\partial J/J} = -\frac{1}{2}$$

$$\frac{\partial MDE/MDE}{\partial m/m} = -\frac{1}{2}\left[\frac{\frac{\sigma_\omega^2}{m} - \frac{\psi^B}{m} - (m-1)\frac{\partial \psi^B}{\partial m} + 2m\frac{\partial \psi^X}{\partial m}}{\left(\frac{m+r}{mr}\right)\sigma_\omega^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X}\right]$$

$$\frac{\partial MDE/MDE}{\partial r/r} = -\frac{1}{2}\left[\frac{\frac{\sigma_\omega^2}{r} - \frac{\psi^A}{r} - (r-1)\frac{\partial \psi^A}{\partial r} + 2r\frac{\partial \psi^X}{\partial r}}{\left(\frac{m+r}{mr}\right)\sigma_\omega^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X}\right]$$

There is a constant elasticity of MDE with respect to J of -0.5, meaning that a 1 percent increase in the number of units always yields a 0.5 percent reduction in the MDE. However, the elasticity of MDE with respect to m and r depends on the error structure and panel length.[38] For some parameter values, this elasticity can be positive, such that increasing the length of the experiment would actually *increase* the MDE. This may seem counter-intuitive, but adding time periods can reduce the average covariance between pre- and post-treatment observations (ψ^X), which introduces more noise in the estimation of pre- vs. post-treatment difference. For relatively short panels with errors that exhibit strong serial correlation, this effect can dominate the benefits of collecting more time periods.

Figure 3.10 illustrates how adding time periods may either increase or decrease power. The left panel plots the MDE of an experiment as a function of the number of pre- and post-treatment periods, holding the number of units J constant. The right panel depicts the tradeoff between adding units vs. time periods by plotting the combinations of J and $m = r$ that yield a given MDE. We use the SCR formula to analytically construct these curve, assuming AR(1) idiosyncratic errors with varying γ values.

At low to moderate levels of serial correlation, increasing the panel length always reduces MDE given J, and vice versa. However, at higher levels of serial correlation, this relationship

[37] Researchers may also adjust P to make an experimental design more cost effective. An RCT will have the lowest MDE at $P = 0.5$, but if control units are cheap compared to treatment units, the same power may be achieved at lower cost by decreasing P and increasing J. See Duflo, Glennerster, and Kremer (2007) for more details. We also typically consider α and κ to be fixed "by convention." While α is the product of research norms, and therefore relatively inflexible, researchers may want to adjust κ. $1 - \kappa$ is the probability of being unable to distinguish a true effect from 0. In lab experiments which are cheaply replicated, researchers may accept $\kappa < 0.80$, whereas in large, expensive field experiments that can only be conducted once, researchers may instead wish to set $\kappa > 0.80$. Researchers may also choose to size their experiments such that they achieve a power of 80 percent for the smallest economically meaningful effect, even if they expect the true MDE to be larger.

[38] Note that J, m, and r must all be integer-valued, hence these derivatives serve as continuous approximations of discrete changes in these parameters. Likewise, the partial derivatives of ψ^B, ψ^A, and ψ^X with respect to m and r are not technically defined, as these covariance terms are averaged over discrete numbers of periods (as shown in Assumption 5).

is no longer monotonic. For $\gamma \geq 0.6$, marginally increasing m or r in a relatively short panel increases MDE for a given J, and likewise increases J required to achieve a given MDE. This suggests that in settings with strong non-constant serial correlation, adding periods of data might *decrease* statistical power if the panel is not sufficiently long.[39]

3.4.2 Power calculations in STATA

To facilitate implementation of *ex ante* power calculations in practice, we have developed the STATA package `pcpanel`.[40] This software implements panel data power calculations that properly account for serial correlation, a feature not present in STATA's built-in `power` command.[41] `pcpanel` operationalizes our SCR formula based on *either* user-input assumptions on the correlation structure *or* nonparametric estimates of variance/covariance parameters from a pre-existing dataset. Our package also executes power calculations by simulation, accommodating a range of experimental designs outside the scope of this paper's analytical framework.

3.4.2.1 Analytical DD power calculations

The program `pc_dd_analytic` conducts analytic power calculations for DD experiments using our SCR formula (Equation (3.2)). `pc_dd_analytic` takes (exactly two of) sample size (J), a desired MDE, and a desired power (κ), and returns the third. It behaves similarly to the legacy STATA function `sampsi`, except that `pc_dd_analytic` accepts the *idiosyncratic* residual variance σ^2_ω, equal to $\sigma^2_\epsilon(1-\rho)$ in the notation of McKenzie (2012).[42]

Users have two options for incorporating non-constant serial correlation. First, they may allow the subprogram `pc_dd_covar` to nonparametrically estimate the average covariance structure of a pre-existing dataset, using the same procedure that generated the solid lines in Figures 3.6 and 3.8. By characterizing the relevant components of the data's actual correlation structure, this enables users with representative pre-existing data to perform *ex ante* power calculations. Second, in the absence of pre-existing data, users may instead input

[39] McKenzie (2012) argues that stronger unit-specific shocks (i.e. higher σ^2_υ) can erode the benefits of collecting additional waves of data. Here, we extend that argument to within-unit serial correlation, demonstrating that higher autocorrelation in the idiosyncratic error term can similarly erode—and even reverse—the benefits of increased panel length. This result reflects the analytical properties of estimators that leverage both pre- and post-treatment data, and does not reflect the DD estimator over-controlling for pre-period data. Appendix Figure B.4 replicates this figure using the SCR ANCOVA formula and finds the same result.

[40] `pcpanel` is available for download from SSC. Both Appendix B.4.4 and the `pcpanel` help file describe the software package in further detail.

[41] While `power` does allow for repeated-measures ANOVA, it cannot accommodate standard panel research designs used in economics, including DD and ANCOVA. The previous STATA power calculations command, `sampsi`, did allow for this, but it was depreciated as of 2013 and is no longer supported.

[42] For example, `-sampsi 0 10, n1(150) n2(150) pre(3) post(5) sd(50) r1(0.3) m(change)-` becomes `-pc_dd_analytic, mde(10) n(300) p(0.5) pre(3) post(5) var(1750)-`, as $1750 = 50^2 * (1-0.3)$. Both of these commands yield a power of 0.81. For very small samples, `sampsi` and `pc_dd_analytic` will yield slightly different results, as `pc_dd_analytic` uses a finite-sample t distribution to calculate critical values, whille `sampsi` uses a Normal distribution.

assumed within-unit AR(1) correlations (γ), assumed average idiosyncratic covariances (ψ^B, ψ^A, ψ^X), or assumed average idiosyncratic correlations (ψ^B/σ^2_ω, ψ^A/σ^2_ω, ψ^X/σ^2_ω).[43]

3.4.2.2 Power calculations by simulation

Simulation-based power calculations are the most robust, flexible strategy for designing experiments *ex ante*. Hence, `pcpanel` includes the program `pc_simulate`, which enables researchers to implement our simulation approach from Section 3.3.[44] `pc_simulate` recovers the *ex ante* power of a given experimental design and MDE—by simulating the *ex post* estimating equation on subsamples of a pre-existing dataset.[45] `pc_simulate` enables power calculations via simulation for four standard treatment effect estimators: cross-sectional ("one-shot"), repeated cross-sections (post-treatment periods only), DD, and ANCOVA. Users may flexibly condition on pre-determined covariates (e.g. household size) or more detailed fixed effects (e.g. unit interacted with month-of-year), both common strategies to increase an experiment's power. `pc_simulate` also accommodates datasets in which a longer panel has been collapsed to a single pre-treatment and post-treatment period, as recommended by Bertrand, Duflo, and Mullainathan (2004) as a way of controlling the false rejection rate.[46]

Finally, `pc_simulate` supports two additional randomization techniques: stratified randomization and cluster randomization. For stratified randomization, users identify one or multiple categorical covariates (e.g. gender), and `pc_simulate` selects PJ treated units and $(1-P)J$ control units within each stratification group. This randomization approach ensures balance across the treated and control groups. For cluster randomization, users input a group identifier (e.g. village) and `pc_simulate` randomizes *groups* of units into treatment (e.g. selecting treated villages, rather than treated households). Users may alter the number of units per group, and also vary the intensity of unit-level treatment within treated groups.[47]

[43] The option `-depvar(y)-` will nonparametrically estimate the average covariance structure of the variable `y`, given the number of pre- and post-treatment periods. Alternatively, adding the option `-ar1(0.4)-` to `-pc_dd_analytic, mde(10) n(300) p(0.5) pre(3) post(5) var(1750)-` reduces power from 0.81 to 0.64. Figures 3.6 and 3.8 illustrate how AR(1) imperfectly approximates the correlation structures present in real data. However, we provide this option for `pc_dd_analytic` to let users without pre-existing data examine the sensitivity of their power calculations to differing levels of serial correlation.

[44] To construct the solid lines in Figures 3.6 and 3.8, we perform *both* analytical and simulation-based power calculations. For each panel length, we calculate the MDE that achieves 80 percent power, based on nonparametric estimates of the covariance structure (as in, run `pc_dd_analytic`, calling subprogram `pc_dd_covar`). Then, we simulate (as in, run `pc_simulate`) to confirm that realized power is indeed 80 percent.

[45] Each iteration re-randomizes PJ units into treatment and adds MDE to treated units' outcomes for all post-treatment periods. Appendix B.4.2 outlines this simulation algorithm in more detail. Users may also provide simulated data based on a plausible DGP. Appendix B.4.3 provides guidance on conducting simulation-based power calculations in the absence of a representative pre-existing dataset.

[46] Collapsing to cross-sectional unit-specific averages (or two-period pre/post averages for DD estimators) obviates the need to adjust the Type I error rate using the CRVE. However, collapsing data does *not* obviate the need to acocunt for serial correlation in power calculations (see Appendix B.1.2.4).

[47] `pc_simulate` does not allow for treatment spillovers, which would require (essentially arbitrary) *ex ante* assumptions on the strength and direction of these spillovers, which are beyond the scope of this work. Researchers interested in randomized saturation designs should consult Baird et al. (2018), which considers these designs in the cross-section. Appendix B.3.3 uses `pc_simulate` to compare cluster randomized designs

3.5 Conclusion

Randomized experiments are costly, and researchers should avoid both underpowered experiments that are uninformative and overpowered experiments that waste resources. *Ex ante* power calculations enable researchers to design experiments with sample sizes that are sufficient, but not excessive. As multi-wave data collection becomes cheaper, panel RCTs are becoming increasingly common. Temporally disaggregated data allow researchers to ask new questions, applying a wider range of empirical methods (McKenzie 2012).

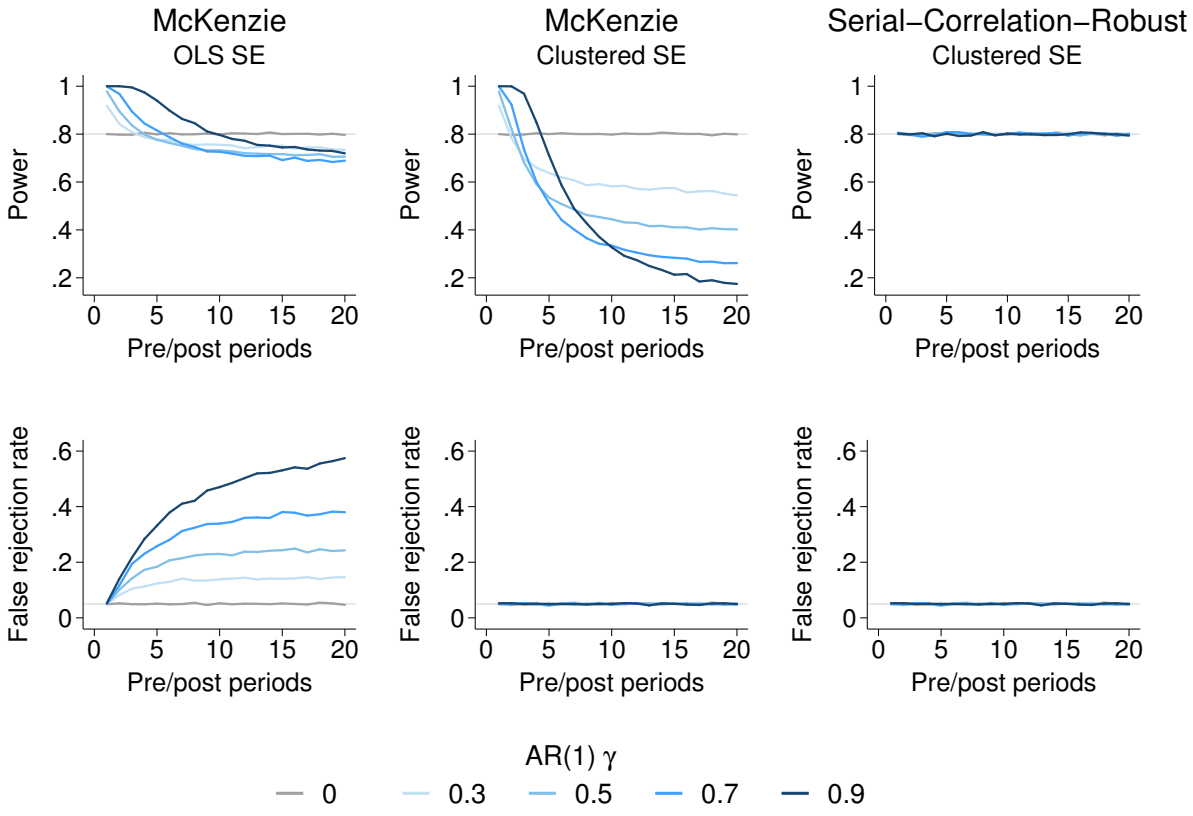
In this paper, we develop a framework for panel data power calculations, generalizing the standard power calculation formula for difference-in-differences (originally derived by Frison and Pocock (1992)) to accommodate fully arbitrary correlations in the error structure. Unlike the power calculation formulas highlighted in McKenzie (2012), our "serial-correlation-robust" formula achieves the desired power in settings with arbitrary serial correlation. These results hold in Monte Carlo simulations, real data from a panel RCT in China, and household electricity consumption data similar to that used in panel RCTs in the energy economics literature.

Our new method is robust to alternative difference-in-differences estimators, to ANCOVA with deterministic time shocks, and to real-world data generating processes. We also provide a framework for trading off minimum detectable effects vs. sample sizes, while discussing practical issues in designing panel RCTs. Our accompanying Stata package `pcpanel` executes both analytical and simulation-based power calculations, and we recommend simulation-based methods when researchers have access to representative pre-existing data *ex ante*. A productive avenue for future work would be extending the treatment interference framework in Baird et al. (2018) to panel data experiments.

to unit-level randomization in fake data. We find that, with 100 percent treatment saturation, cluster randomized DD RCTs achieve the same power as DD RCTs with the same number of units but randomized at the unit level. We recommend that users interested in cluster randomized designs use `pc_simulate`, to account for complexities that may arise in real data.

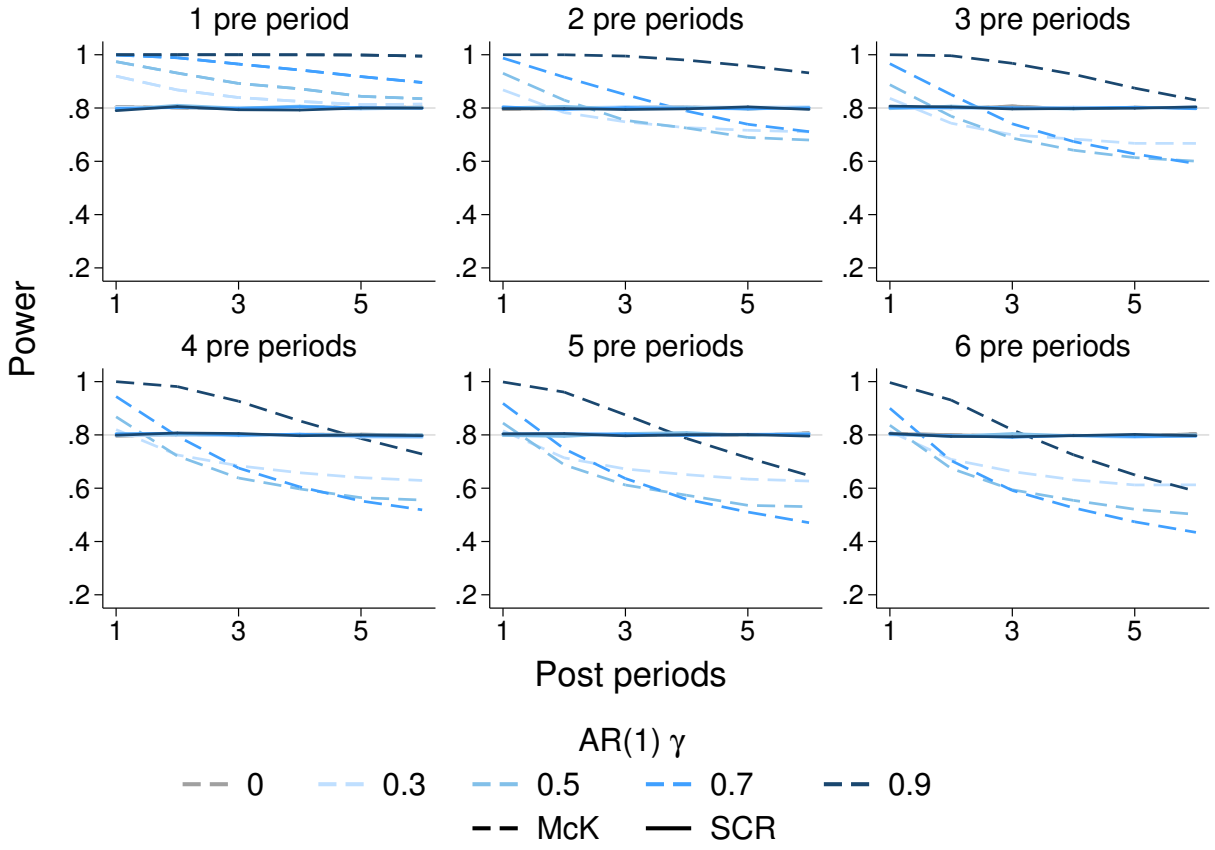
Figures

Figure 3.1: Standard methods result in improperly powered experiments in AR(1) data



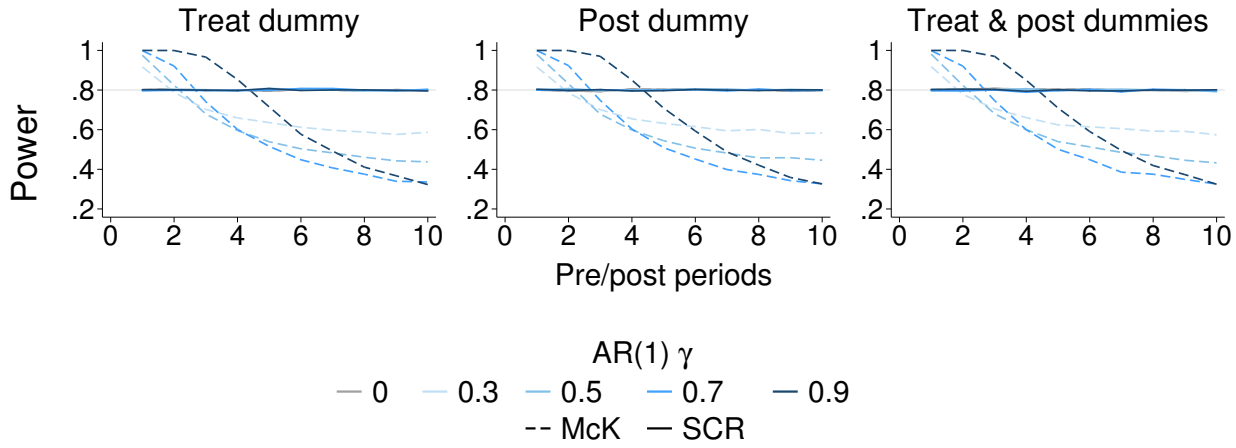
Notes: This figure displays power and false rejection rates from performing power calculations with three different sets of assumptions, using simulated data with AR(1) processes (with differing levels of serial correlation) and differing panel lengths (ranging from 2 ($m = r = 1$) to 40 ($m = r = 20$) periods). In the left column, we apply the standard McKenzie formula (Equation (3.3), which assumes away non-constant serial correlation), and use OLS standard errors *ex post*, in line with the assumptions of this formula. In the middle column, we again apply the McKenzie formula, but cluster standard errors *ex post*—which is inconsistent with the *ex ante* formula, but corrects for within-unit serial correlation following Bertrand, Duflo, and Mullainathan (2004). In the right column, we apply our serial-correlation-robust formula (Equation (3.2), which accounts for non-i.i.d. errors *ex ante*), and cluster standard errors *ex post*. As expected, this third set of simulations achieves the desired 80 percent power and 5 percent false rejection rate.

Figure 3.2: Power in short panels – AR(1) data



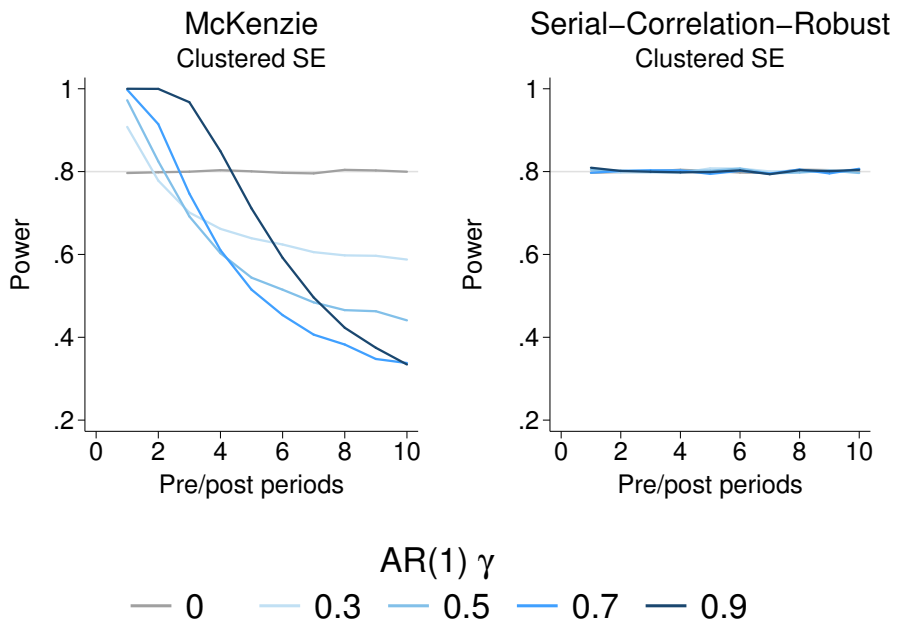
Notes: This figure repeats the same simulation exercise as Figure 3.1, except that we separately vary the number of pre-treatement and post-treatment periods. Each panel conducts simulations with a distinct number of pre-treatment periods ($1 \leq m \leq 6$), and each horizontal axis varies the number of post-treatment periods ($1 \leq r \leq 6$). Dotted lines report realized power for experiments calibrated using the McKenzie formula (i.e. the top-middle panel of Figure 3.1), while solid lines report realized power using the SCR formula (i.e. the top-right panel of Figure 3.1). For all cases where $m = 1$ or $r = 1$, the McKenzie formula yields over-powered experiments across the full range of positive AR(1) parameters. This shows that traditional "one baseline, one follow-up" experiments will calibrate to excessively large sample sizes if they ignore non-constant serial correlation *ex ante*. The SCR formula is properly powered in all cases.

Figure 3.3: SCR outperforms traditional methods with alternative estimators



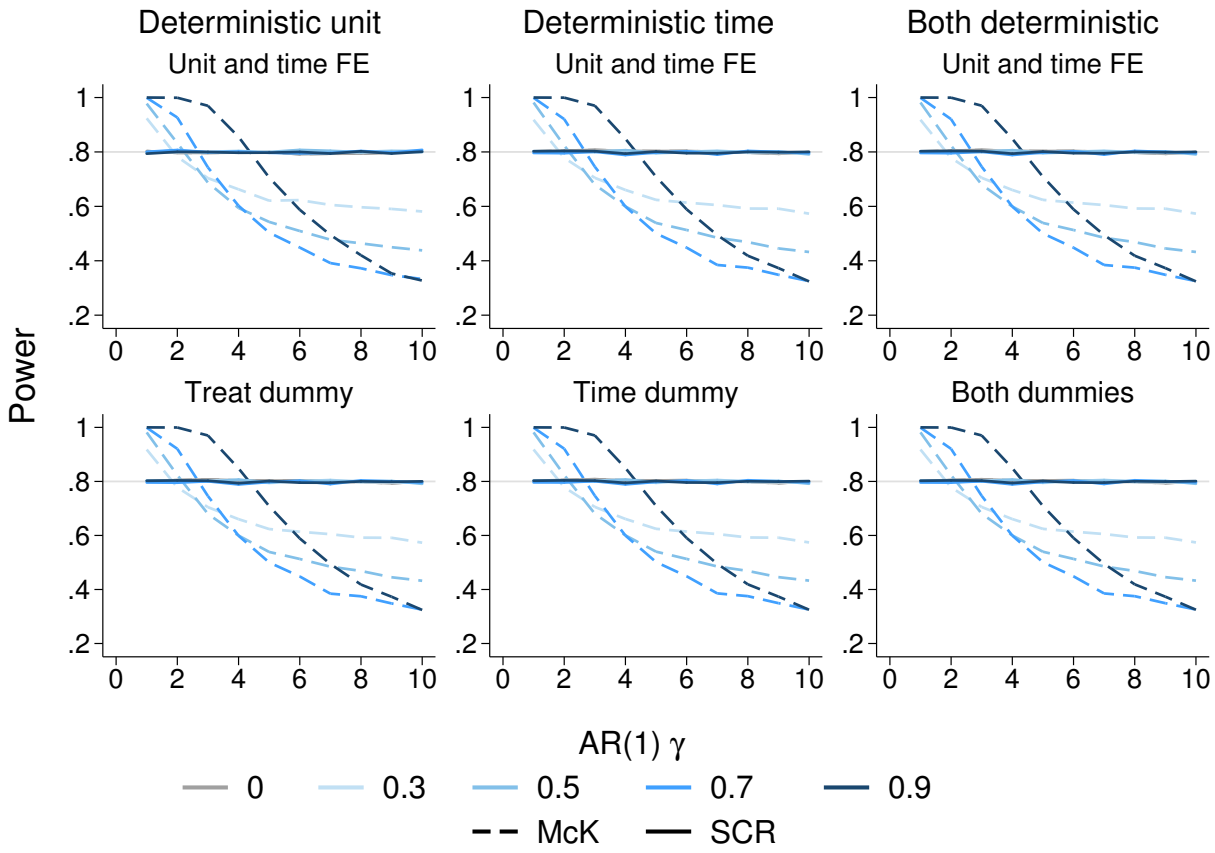
Notes: This figure repeats the same simulation exercise as Figure 3.1, except that each panel alters the *ex post* DD estimating equation. The left panel replaces unit fixed effects with a Treat_i dummy. The middle panel replaces time fixed effects with a Post_t dummy. The right panel replaces both sets of fixed effects with Treat_i and Post_t dummies. Dotted lines report realized power for experiments calibrated using the McKenzie formula (i.e. the top-middle panel of Figure 3.1), while solid lines report realized power using the SCR formula (i.e. the top-right panel of Figure 3.1). All three DD estimating equations yield power that is identical to Figure 3.1 (estimated with unit and time fixed effects). This is because the variance of these estimators is a function of the underlying error structure, and the CRVE correclty accounts for unmodeled serial correlation in each case. Hence, the performance gap between McKenzie vs. SCR formulas remains for all DD estimators, provided that researchers account for serial correlation when conducting *ex post* inference.

Figure 3.4: Standard ANCOVA methods fail under serial correlation



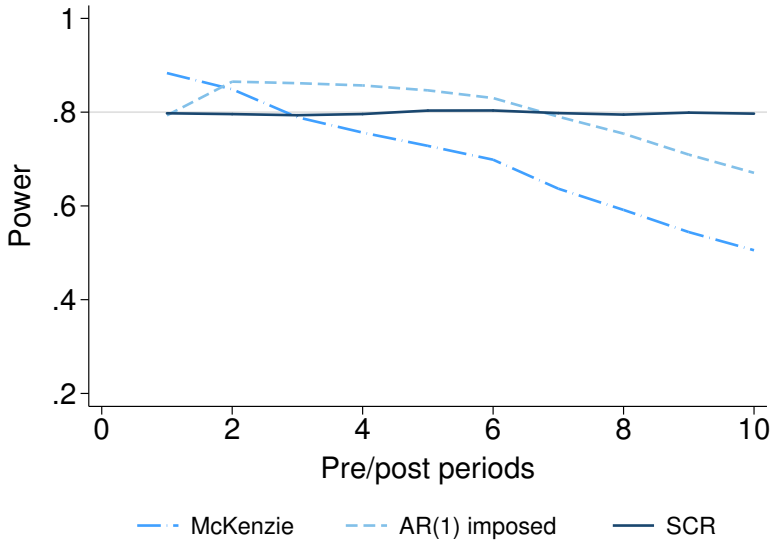
Notes: This figure repeats the same simulation exercise as Figure 3.1, using ANCOVA power calculation formulas *ex ante* and an ANCOVA estiamting equation *ex post*. The left panel calibrates MDE using the McKenzie ANCOVA formula (Equation (3.10), and is analogous to the top-middle panel of Figure 3.1. The right panel calibrates MDE using the SCR ANCOVA formula (Equation (3.11), and is analogous to the top-righte panel of Figure 3.1. Both set of simulations estimate Equation (3.9) *ex post*, cluster standard errors by unit, and simulate DGPs without random time shocks ($\sigma^2_\delta = 0$). As in the DD simulations, the McKenzie ANCOVA formula fails to generate correctly-powered experiments, whereas the SCR formula is properly powered across all panel lengths and degrees of AR(1) correlation.

Figure 3.5: SCR outperforms traditional methods with alternative DGPs



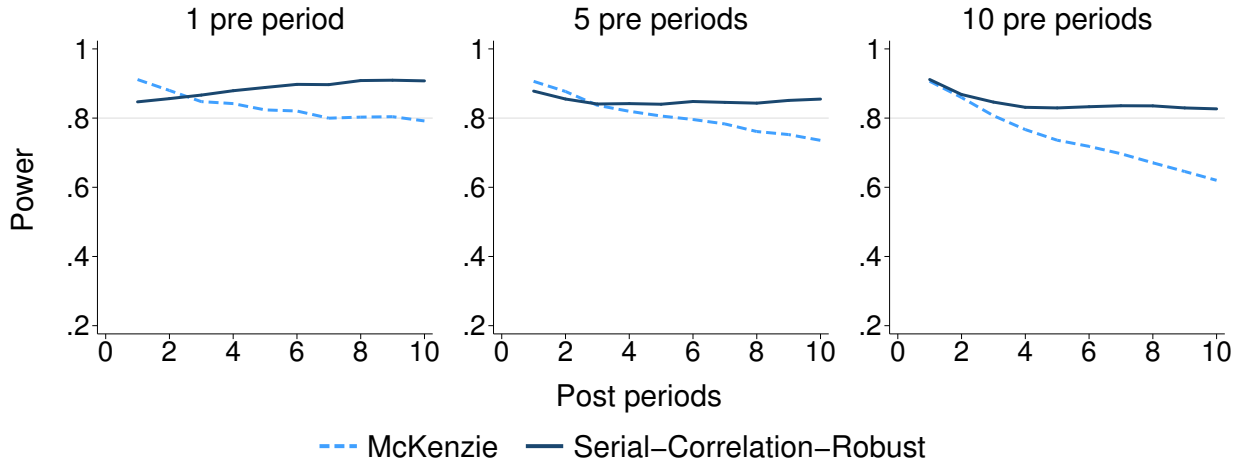
Notes: This figure repeats the same simulation exercise as Figure 3.1, for alternative DGPs and DD estimating equations. The left column removes random unit effects from the DGP ($\sigma_v^2 = 0$); the middle column removes random time effects from the DGP ($\sigma_\delta^2 = 0$); and the right column removes both random unit effects and random time effects ($\sigma_v^2 = \sigma_\delta^2 = 0$). The top row uses an *ex post* estimator with unit and time fixed effects (as in Figure 3.1, albiet now misspecified); the bottom rows replaces fixed effects with dummy variables (Treat_i and/or Post_t, as in Figure 3.3) to align with each DGP, respectively. Dotted lines report realized power for experiments calibrated using the McKenzie formula (i.e. the top-middle panel of Figure 3.1), while solid lines report realized power using the SCR formula (i.e. the top-right panel of Figure 3.1). In all cases, our results are identical to Figure 3.1 (DGP with random unit and time effects; DD estimator with unit and time fixed effects). This show that our SCR formula is robust to alternative DGPs, which may not match the DD estimator. By contrast, the McKenzie formula continues to perform poorly in the presence of non-constant serial correlation—even after removing intracluster correlation, random time shocks, and fixed effects.

Figure 3.6: Power simulations for Bloom et al. (2015) data



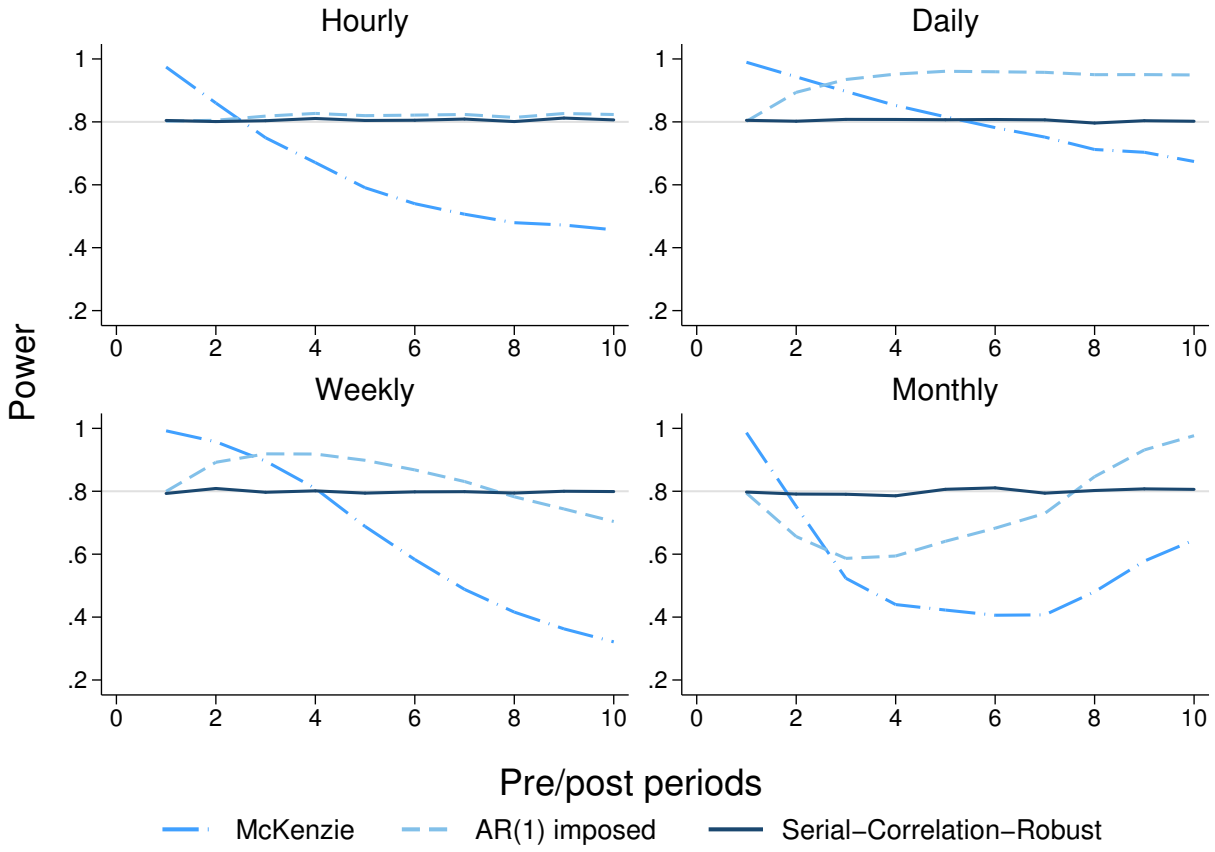
Notes: This figure shows results from Monte Carlo simulations using Bloom et al. (2015) data. Each curve plots realized power for a DD experiment with a certain number of pre/post-treatment periods (from $m = r = 1$ to $m = r = 10$), using different *ex ante* assumptions. The long-dashed line applies the McKenzie formula (Equation (3.3)), which assumes away non-constant serial correlation. The short-dashed line applies the SCR formula, under the assumption of AR(1) serial correlation (using Equation (3.6) to estimate an AR(1) parameter). The solid line applies the SCR formula (Equation (3.2)), where we non-parametrically estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ terms from the (residualized) Bloom et al. (2015) dataset. All simulations apply the CRVE *ex post*, clustering at the individual level. Only the SCR formula achieves the desired 80 percent power, even though the Bloom et al. (2015) data exhibit relatively weak serial correlation.

Figure 3.7: ANCOVA power calculations in real data



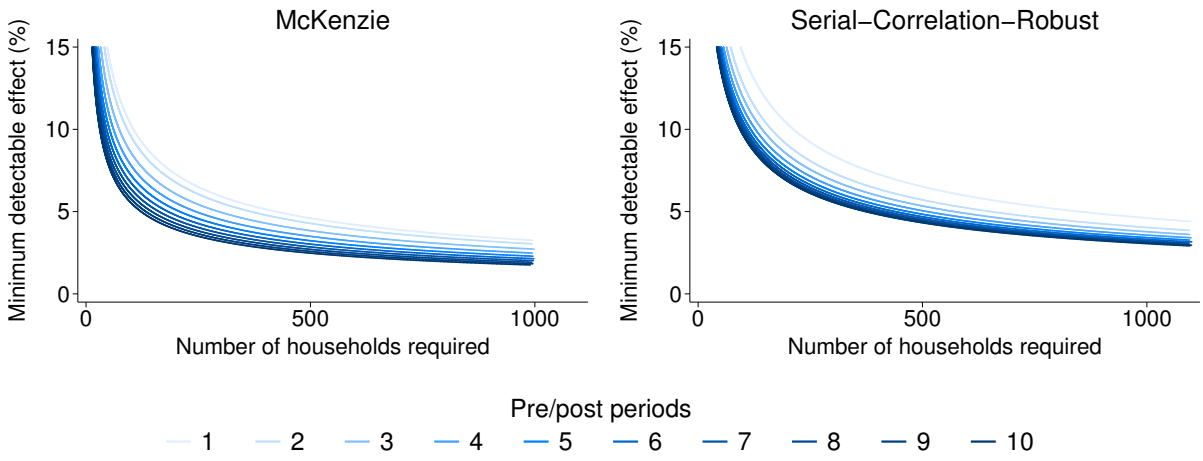
Notes: This figure repeats the same exercise as Figure 3.6, for ANCOVA experiments simulated on the Bloom et al. (2015) dataset. Each panel simulates experiments with a certain number of pre-treatment periods ($m \in \{1, 5, 10\}$), and horizontal axes vary the number of post-treatment periods ($1 \leq r \leq 10$). Dotted lines apply the McKenzie ANCOVA formula *ex ante* (Equation (3.10)), which assumes away non-constant serial correlation. Solid lines apply the SCR ANCOVA formula *ex ante* (Equation (3.11)), where we non-parametrically estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ terms from the (residualized) Bloom et al. (2015) dataset. All cases estimate ANCOVA *ex post* (Equation (3.9)), clustering standard errors by individual. Both formulas generate improperly powered experiments with real data that likely have time shocks, which neither formula can account for.

Figure 3.8: Power simulations for Pecan Street data



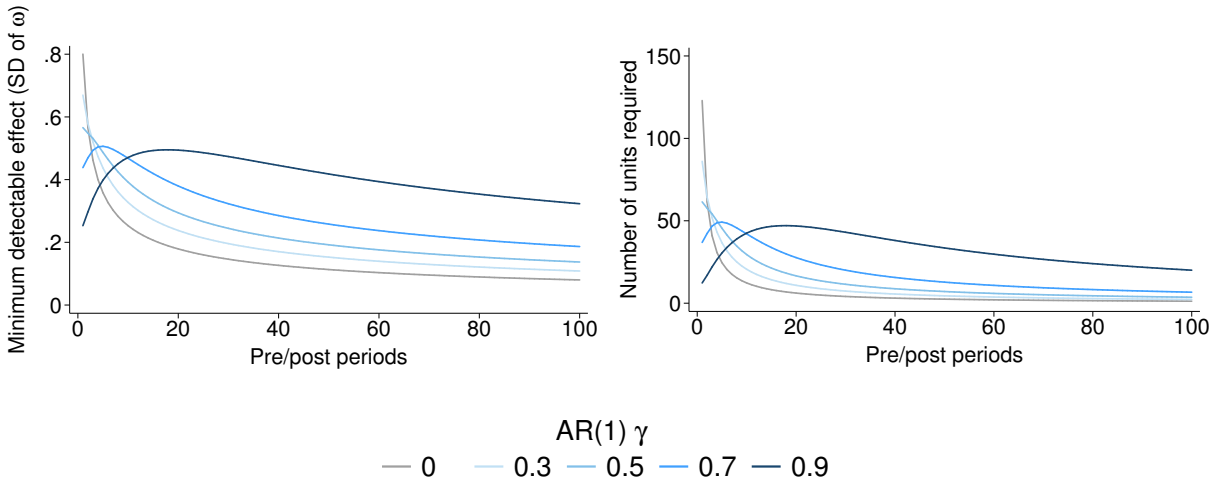
Notes: This figure conducts simulations that are identical to Figure 3.6, using four Pecan Street datasets collapsed to different levels of temporal frequency. Each curve plots realized power for a DD experiment with a certain number of pre/post-treatment periods (from $m = r = 1$ to $m = r = 10$), using different *ex ante* assumptions. The long-dashed lines apply the McKenzie formula (Equation (3.3)), which assumes away non-constant serial correlation. The short-dashed lines apply the SCR formula, under the assumption of AR(1) serial correlation (using Equation (3.6) to estimate an AR(1) parameter). The solid lines apply the SCR formula (Equation (3.2)), where we non-parametrically estimate $\psi^B_{\tilde{\omega}}$, $\psi^A_{\tilde{\omega}}$, and $\psi^X_{\tilde{\omega}}$ terms from the (residualized) Pecan Street data. All simulations apply the CRVE *ex post*, clustering at the individual level. While each temporal frequencies exhibits a unique correlation structure, only the SCR power calculation formula achieves the desired power in each case.

Figure 3.9: Analytical power calculations – daily Pecan Street dataset



Notes: This figure shows the result of analytic power calculations on Pecan Street electricity data, collapsed to the daily level. Each curve displays the number of units required to detect a given minimum detectable effect with 80 percent power. Each iso-power curve corresponds to a particular panel length, with the shortest panel (1 pre-period, 1 post-period) in light blue, and the longest panel (10 pre-periods, 10 post-periods) in navy. The left panel shows a power calculation using the standard McKenzie (2012) formula, which does not accommodate non-constant serial correlation. The right panel applies the serial-correlation-robust formula, which accounts for the real error structure of the data. In this setting, failing to account for the full covariance structure will dramatically understate the sample size required to detect a given effect.

Figure 3.10: Analytical power calculations with increasing panel length



Notes: This figure displays the results of analytical power calculations using the SCR formula for varying AR(1) parameters. The left panel shows the tradeoff between the minimum detectable effect (MDE) and the number of time periods ($m = r$) for varying levels of serial correlation, holding the number of units fixed at $J = 100$ and normalizing MDE by the standard deviation of ω_{it}. At low levels of γ, MDE declines monotonically in m and r. However, for higher γ, increasing m and r actually *increases* MDE when $m = r$ is relatively small, and *decreases* MDE when $m = r$ is relatively large. The right panel shows the relationship between the number of units (J) and number of pre/post periods ($m = r$) required to detect an MDE equal to one standard deviation of ω_{it}. Similarly, for low levels of serial correlation, the trade-off between J and $m = r$ is monotonic. However, as γ increases, adding periods in short panels necessitates a greater number of units to achieve the same MDE, while adding periods in longer panels means that fewer units are required to achieve the same MDE. Appendix Figure B.4 replicates this result using the SCR ANCOVA formula.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" ArXiv Working Paper No. 1710.02926.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and Innovation: An Inverted-U Relationship." *The Quarterly Journal of Economics* 120 (2): 701–728.

Alberini, Anna, and Massimo Filippini. 2011. "Response of residential electricity demand to price: The effect of measurement error." *Energy Economics* 33:889–895.

Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics* 95 (9): 1082–1095.

Allcott, Hunt, and Michael Greenstone. 2017. "Measuring the Welfare Effects of Residential Energy Efficiency Programs." National Bureau of Economic Research Working Paper No. 23386.

Arellano, Manuel. 1987. "Computing Robust Standard Errors for Within-Group Estimators." *Oxford Bulletin of Economics and Statistics* 49 (4): 431–34.

Athey, Susan, and Guido Imbens. 2017a. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31 (2): 3–32.

Athey, Susan, and Guido W. Imbens. 2017b. "The Econometrics of Randomized Experiments." *Handbook of Economic Field Experiments* 1:73–140.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen. 2017. "Organization Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan." *The Quarterly Journal of Economics* 132 (3): 1101–1164.

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics* 132 (2): 551–615.

Auffhammer, Maximilian. 2018. "Climate Adaptive Response Estimation: Short and Long Run Impacts of Climate Change on Residential Electricity and Natural Gas Consumption Using Big Data." NBER Working Paper 24397.

Auffhammer, Maximilian, Patrick Baylis, and Catherine H. Hausman. 2017. "Climate Change Is Projected to Have Severe Impacts on the Frequency and Intensity of Peak Electricity Demand Across the United States." *Proceedings of the National Academy of Sciences* 114 (8): 1886–1891.

Auffhammer, Maximilian, and Erin T. Mansur. 2014. "Measuring Climatic Impacts on Energy Consumption: A Review of the Empirical Literature." *Energy Economics* 46:522–530.

Badiani, Reena, and Katrina Jessoe. 2015. "The impact of electricity subsidies on groundwater extraction and agricultural production." Working paper.

Baerenklau, Kenneth A., Kurt A. Schwabe, and Ariel Dinar. 2014. "The residential water demand effect of increasing block rate water budgets." *Land Economics* 90 (4): 683–699.

Bain, Joe S. 1951. "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936–1940." *The Quarterly Journal of Economics* 65 (3): 293–324.

Baird, Sarah, J. Aislinn Bohren, Craig McIntosh, and Berk Özler. 2018. "Optimal Design of Experiments in the Presence of Interference." *Review of Economics and Statistics* 100 (5): 844–860.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain." *Econometrica* 80 (6): 2369–2429.

Bergquist, Lauren Falcao. 2017. "Pass-Through, Competition, and Entry in Agricultural Markets: Experimental Evidence from Kenya." Working Paper.

Berry, Steven, and Peter Reiss. 2007. "Empirical Models of Entry and Market Structure." In *Handbook of Industrial Organization: Volume 3,* edited by Mark Armstrong and Robert H. Porter, 1845–1886. Amsterdam: Elsevier, North Holland.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-In-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–275.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. "Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda." *The Quarterly Journal of Economics* 129 (2): 697–752.

Blonz, Joshua A. 2016. "Making the best of the second-best: Welfare consequences of time-varying electricity prices." Energy Institute at Haas Working Paper 275.

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs." *Evaluation Review* 19 (5): 547–556.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1): 1–51.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *The Quarterly Journal of Economics* 130 (1): 165–218.

Borenstein, Severin, and James Bushnell. 1999. "An Empirical Analysis of the Potential for Market Power in California's Electricity Industry." *The Journal of Industrial Economics* 47 (3): 285–323.

Borenstein, Severin, James B. Bushnell, and Frank A. Wolak. 2002. "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market." *American Economic Review* 92 (5): 1376–1405.

Borenstein, Severin, James Bushnell, and Steven Stoft. 2000. "The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry." *The RAND Journal of Economics* 31 (2): 294–325.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization: Volume 2,* edited by Richard Schmalensee and Robert Willig, 1011–1057. Amsterdam: Elsevier, North Holland.

Bresnahan, Timothy F., and Peter C. Reiss. 1991. "Entry and Competition in Concentrated Markets." *Journal of Political Economy* 99 (5): 977–1009.

Bruno, Ellen, and Katrina Jessoe. 2018. "Water markets and climate change adaptation: Micro-level evidence on agricultural water demand." Mimeo.

Buck, Steven, Maximilian Auffhammer, Stephen Hamilton, and David Sunding. 2016. "Measuring welfare losses from urban water supply disruptions." *Journal of the Association of Environmental and Resource Economists* 3 (3): 743–778.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. "Global Non-Linear Effect of Temperature on Economic Production." *Nature* 527 (7577): 235–239.

Bushnell, James B., Erin T. Mansur, and Celeste Saravia. 2008. "Vertical Arrangements, Market Structure, and Competition: An Analysis of Restructured US Electricity Markets." *American Economic Review* 98 (1): 237–266.

Bushnell, James B., and Catherine Wolfram. 2005. "Ownership Change, Incentives and Plant Efficiency: The Divestiture of U.S. Electric Generation Plants." Center for the Study of Energy Markets Working Paper 140.

California Department of Food and Agriculture. 2011. "California agricultural resource directory 2010–2011."

California Energy Commission. 2005. "California's water-energy relationship." CEC Report 700-2005-011-SF.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50 (2): 317–372.

Campbell, Cathy. 1977. "Properties of Ordinary and Weighted Least Square Estimators of Regression Coefficients for Two-Stage Samples." *Proceedings of the Social Statistics Section, American Statistical Association:* 800–805.

Card, David, Stefano DellaVigna, and Ulrike Malmendier. 2011. "The Role of Theory in Field Experiments." *Journal of Economic Perspectives* 25 (3): 39–62.

Cohen, Jacob. 1977. *Statistical Power Analysis for the Behavioral Sciences.* New York, NY: Academic Press.

Cournot, Augustin. 1838. "Of the Influence of Taxation on Commodities Produced under a Monopoly." In *Researches into the Mathematical Principles of the Theory of Wealth,* translated by Nathaniel T. Bacon, 67–78. New York: The Macmillan Company, 1897.

Cramton, Peter. 2017. "Electricity Market Design." *Oxford Review of Economic Policy* 33 (4): 589–612.

Davis, Lucas, and Catherine Hausman. 2016. "Market Impacts of a Nuclear Power Plant Closure." *American Economic Journal: Applied Economics* 8 (2): 92–122.

De Loecker, Jan, and Jan Eeckhout. 2017. "The Rise of Market Power and the Macroeconomic Implications." NBER Working Paper 23687.

———. 2018. "Global Market Power." NBER Working Paper 24768.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics* 4 (3): 66–95.

Deryugina, Tatyana, Alexander MacKay, and Julian Reif. 2017. "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation." NBER Working Paper 23483.

Deschênes, Olivier, and Michael Greenstone. 2007. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *American Economic Review* 97 (1): 354–385.

———. 2011. "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US." *American Economic Journal: Applied Economics* 3 (4): 152–185.

Deutch, John M., Charles W. Forsberg, Andrew C. Kadak, Mujid S. Kazimi, Ernest J. Moniz, and John E. Parsons. 2009. "Update of the MIT 2003 Future of Nuclear: An Interdisciplinary MIT Study."

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." Chap. 61 in *Handbook of Development Economics,* edited by Paul T. Schultz and John A. Strauss, 3895–3962. Volume 4. Oxford, UK: Elsevier.

ERCOT (Electricity Reliability Council of Texas). 2018. "ERCOT Quick Facts." August 8, 2018.

Fabra, Natalia, and Mar Reguant. 2014. "Pass-Through of Emissions Costs in Electricity Markets." *American Economic Review* 104 (9): 2872–2899.

Fell, Harrison, Shanjun Li, and Anthony Paul. 2014. "A new look at residential electricity demand using household expenditure data." *International Journal of Industrial Organization* 33:37–47.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2018. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." *The Quarterly Journal of Economics* 133 (3): 1597–1664.

Fowlie, Meredith, Catherine Wolfram, C. Anna Spurlock, Annika Todd, Patrick Baylis, and Peter Cappers. 2017. "Default Effects and Follow-on Behavior: Evidence from an Electricity Pricing Program." National Bureau of Economic Research Working Paper No. 23553.

Frison, L., and S. J. Pocock. 1992. "Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and its Implications for Design." *Statistics in Medicine* 11 (13): 1685–1704. pmid: 1485053.

Gaynor, Martin, and Robert J. Town. 2011. "Competition in Health Care Markets." In *Handbook of Health Economics: Volume 2,* edited by Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, 499–637. Handbook of Health Economics. Amsterdam: Elsevier, North Holland.

Glennerster, Rachel, and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide.* Princeton University Press.

Green, Richard J., and David M. Newbery. 1992. "Competition in the British Electricity Spot Market." *Journal of Political Economy* 100 (5): 929–953.

Hagerty, Nick. 2018. "Liquid constrained in California: Estimating the potential gains from water markets." Working paper.

Harberger, Arnold C. 1954. "Monopoly and Resource Allocation." *The American Economic Review* 44 (2): 77–87.

Hendricks, Nathan P., and Jeffrey M. Peterson. 2012. "Fixed effects estimation of the intensive and extensive margins of irrigation water." *Journal of Agricultural and Resource Economics* 37 (1): 1–19.

Hewitt, J.A., and W.M. Hanemann. 1995. "A discrete/continuous choice approach to residential water demand under block rate pricing." *Land Economics* 71 (2): 173–192.

Hortaçsu, Ali, Fernando Luco, Steven L Puller, and Dongni Zhu. 2017. "Does Strategic Ability Affect Efficiency? Evidence from Electricity Markets." NBER Working Paper 23526.

Hortaçsu, Ali, and Steven L. Puller. 2008. "Understanding Strategic Bidding in Multi-Unit Auctions: a Case Study of the Texas Electricity Spot Market." *The RAND Journal of Economics* 39 (1): 86–114.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. 2013. "Quantifying the Influence of Climate on Human Conflict." *Science* 341 (6151): 1235367.

Hsiang, Solomon M., Kyle C. Meng, and Mark A. Cane. 2011. "Civil Conflicts Are Associated with the Global Climate." *Nature* 476 (7361): 438–441.

Ito, Koichiro. 2013. "How do consumers respond to nonlinear pricing? Evidence from household water demand." Working paper.

———. 2014. "Do consumers respond to marginal or average price? Evidence from nonlinear electricity pricing." *American Economic Review* 104 (2): 537–563.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka. 2018. "Moral suasion and economic incentives: Field experimental evidence from electricity demand." *American Economic Journal: Economic Policy* 10 (1): 240–267.

Ito, Koichiro, and Mar Reguant. 2016. "Sequential Markets, Market Power, and Arbitrage." *American Economic Review* 106 (7): 1921–1957.

Jessoe, Katrina, and David Rapson. 2014. "Knowledge is (less) power: Experimental evidence from residential electricity use." *American Economic Review* 104 (4): 1417–1438.

———. 2015. "Commercial and industrial demand response under mandatory time-of-use electricity pricing." *Journal of Industrial Economics* 63 (3): 397–421.

Johnson, Renee, and Betsy A. Cody. 2015. "California agricultural production and irrigated water use." Congressional Research Service report.

Joskow, Paul L., and Jean Tirole. 2000. "Transmission Rights and Market Power on Electric Power Networks." *The RAND Journal of Economics* 31 (3): 450–487.

McDermott, Grant R. 2018. "Hydro Power. Market Might." Working Paper.

McKenzie, David. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–221.

———. 2017. "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition." *American Economic Review* 107 (8): 2278–2307.

Mercadal, Ignacia. 2018. "Dynamic Competition and Arbitrage in Electricity Markets: The Role of Financial Players." Working Paper.

Mieno, Taro, and Nicholas Brozovic. 2017. "Price Elasticity of Groundwater Demand: Attenuation and Amplification Bias Due to Incomplete Information." *American Journal of Agricultural Economics* 99 (2): 401–426.

Moulton, Brent. 1986. "Random group effects and the precision of regression estimates." *Journal of Econometrics* 32 (3): 385–397.

Murphy, Kevin, Brett Myors, and Allen Wolach. 2014. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests.* 4th ed. New York, NY: Routledge.

Nataraj, Shanthi. 2011. "Does marginal price matter? A regression discontinuity approach to estimating water demand." *Journal of Environmental Economics and Management* 61:198–212.

Olmstead, S.M., Michael W. Hanemann, and R.N. Stavins. 2007. "Water demand under alternative price structures." *Journal of Environmental Economics and Management* 54 (2): 181–198.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press.

Paul, Anthony, Erica Myers, and Karen Palmer. 2009. "A partial adjustment model of U.S. electricity demand by region, season, and sector." Resources for the Future Working Paper DP 08-50.

Pecan Street. 2016. "Dataport." `https://dataport.pecanstreet.org/`.

Pfeiffer, Lisa, and C.-Y. Cynthia Lin. 2014. "The effects of energy prices on agricultural groundwater extraction from the High Plains aquifer." *American Journal of Agricultural Economics* 96 (5): 1349–1362.

Potomac Economics. 2015. "2014 State of the Market Report for the ERCOT Wholesale Electricity Markets."

Provencher, Bill, and Oscar Burt. 1993. "The Externalities Associated with the Common Property Exploitation of Groundwater." *Journal of Environmental Economics and Management* 24:139–158.

Puller, Steven L. 2007. "Pricing and Firm Conduct in California's Deregulated Electricity Market." *The Review of Economics and Statistics* 89 (1): 75–87.

Reguant, Mar. 2014. "Complementary Bidding Mechanisms and Startup Costs in Electricity Markets." *The Review of Economic Studies* 81 (4): 1708–1742.

Reiss, Peter C., and Matthew W. White. 2005. "Household electricity demand, revisited." *Review of Economic Studies* 72:853–883.

Renwick, E.M., and R.D. Green. 2000. "Do residential water demand side management policies measure up? An analysis of eight California water agencies." *Journal of Environmental Economics and Management* 40 (1): 37–55.

Riahi, Keywan, Shilpa Rao, Volker Krey, Cheolhung Cho, Vadim Chirkov, Guenther Fischer, Georg Kindermann, Nebojsa Nakicenovic, and Peter Rafaj. 2011. "RCP 8.5—A Scenario of Comparatively High Greenhouse Gas Emissions." *Climatic Change* 109 (1): 33.

Ryan, Nicholas. 2017. "The Competitive Effects of Transmission Infrastructure in the Indian Electricity Market." NBER Working Paper 23106.

Sathaye, Jayant A., Larry L. Dale, Peter H. Larsen, Gary A. Fitts, Kevin Koy, Sarah M. Lewis, and André Frossard Pereira de Lucena. 2013. "Estimating impacts of warming temperatures on California's electricity system." *Global Environmental Change* 23 (2): 499–511.

Sawyers, Gary. 2007. "A primer on California water rights." UC Davis Cooperative Extension working paper.

Schlenker, Wolfram, W. Michael Hanemann, and Anthony C. Fisher. 2005. "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach." *American Economic Review* 95 (1): 395–406.

Schlenker, Wolfram, and Michael J. Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields Under Climate Change." *Proceedings of the National Academy of Sciences* 106 (37): 15594–15598.

Schlenker, Wolfram, and W. Reed Walker. 2016. "Airports, Air Pollution, and Contemporaneous Health." *The Review of Economic Studies* 83 (2): 768–809.

Schoengold, Katrina, David L. Sunding, and Georgina Moreno. 2014. "Price elasticity reconsidered: Panel estimation of an agricultural water demand function." *Water Resources Research* 42.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.

Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg,* edited by Donald W. K. Andrews and James H. Stock. Cambridge, UK: Cambridge University Press.

Sunding, David L., David Zilberman, and Nicholas Brozovic. 2010. "On the spatial nature of the groundwater pumping externality." *Resource and Energy Economics* 32 (2): 154–164.

Teerenstra, Steven, Sandra Eldridge, Maud Graff, Esther de Hoop, and George F. Borm. 2012. "A simple sample size formula for analysis of covariance in cluster randomized trials." *Statistics in Medicine* 31 (20): 2169–2178.

White, Halbert. 1984. *Asymptotic Theory for Econometricians.* 1st ed. San Diego, CA: Academic Press.

Wichman, Casey J. 2014. "Perceived price in residential water demand: Evidence from a natural experiment." *Journal of Economic Behavior and Organization* 107:308–323.

Wichman, Casey J., Laura O. Taylor, and Roger H. von Haefen. 2016. "Conservation policies: Who responds to price and who responds to prescription?" *Journal of Environmental Economics and Management* 79:114–134.

Wilson, Robert. 1979. "Auctions of Shares." *The Quarterly Journal of Economics* 93 (4): 675–689.

Wolak, Frank A. 2011. "Do residential customers respond to hourly prices? Evidence from a dynamic pricing experiment." *American Economic Review* 101 (3): 83–87.

Wolfram, Catherine D. 1998. "Strategic Bidding in a Multiunit Auction: An Empirical Analysis of Bids to Supply Electricity in England and Wales." *The RAND Journal of Economics* 29 (4): 703–725.

———. 1999. "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review* 89 (4): 805–826.

Appendix A

Appendix: Market Size and Market Power: Evidence from the Texas Electricity Market

A.1 Texas electricity market

A.1.1 Bilateral contracts and day-ahead market

A substantial portion of the electricity sold in ERCOT's real-time market is covered by financial positions taken prior to the market clearing.[1] Many firms selling electricity in ERCOT sign bilateral contracts with wholesale purchasers, typically more than a year in advance of the contracted date. These contracts are purely financial and do not require physical delivery of electricity. ERCOT settles each firm's financial position following the real-time market, so no funds need to be directly exchanged between market participants.

The first auction operated by ERCOT is the day-ahead market, which occurs daily at 10 a.m. This auction is voluntary and serves as a centralized forward market for the following day's electricity production. As with bilateral contracts, the day-ahead market outcomes are financial and do not require physical delivery. Electricity producers who want to participate in this market submit an offer[2] for each hour of the following day. Market participants can also submit bids to purchase electricity for each hour of the following day. Both offers to supply and bids to purchase electricity must specify a location on the grid at which the electricity is to be traded. ERCOT clears this auction by maximizing total surplus subject

[1] Hortaçsu and Puller (2008) report only 2–5% of energy was traded in the spot market under ERCOT's previous market design. More recent reports simply state "only a small share of the power produced in ERCOT is transacted in the spot market" (Potomac Economics 2015), referring to the net transactions not covered by existing financial positions.

[2] The offers in the day-ahead market differ from the offer curves in the real-time market, which I discuss in detail in Section 1.2. The day-ahead offers can take the form of a three-part offer, which includes not only an offer curve but also a startup offer and a minimum energy offer. See Reguant (2014) for an analysis of these complex bids in the Spanish electricity market.

to the balance of supply and demand and the physical constraints of the grid. This market-clearing algorithm may result in day-ahead prices varying throughout the ERCOT grid.

The forward positions that result from bilateral contracts and the day-ahead market play an important role in firm strategy in the real-time market. Because these positions are financially binding, they change the incentives of firms to exercise market power in the real-time market and may even affect whether a firm wants to increase or decrease the real-time price. I discuss these incentives and incorporate them into my theoretical model in Section 1.3 and Appendix A.2.

A.1.2 Offer price mitigation

ERCOT recognizes that transmission congestion and local markets have the potential to give a firm an extreme amount of market power. Although ERCOT typically does not attempt to restrict the ability of a firm to exercise market power, it will mitigate offer curves when one firm is effectively a monopolist within a local market. If a constraint fails the "competitive constraint test," which involves complex engineering calculations, all offer curves on the import side of the constraint are mitigated. When mitigated, offer prices are capped at an assumed marginal cost based on an index of current fuel prices, which in turn caps the resulting LMP.[3] Although the mitigated offer cap will affect a firm's ultimate ability to exercise market power, it is unlikely to affect a firm's strategy. Firms cannot easily manipulate the inputs into the competitive constraint test, and even in the event the test fails, the firm has no incentive to bid differently from its non-mitigated optimum. Additionally, only a small fraction of offer curves are mitigated in practice (Potomac Economics 2015).

A.2 Theoretical framework

A.2.1 Optimal markup rule

In this section I derive the optimal markup rule, in Equation (1.3), from the firm optimization problem, in Equation (1.1). I first restate the firm's problem:

$$\max_{\{S_i|i\in\mathcal{I}_f\}} \mathbb{E}\left[\sum_{i\in\mathcal{I}_f}(S_i(p_i)p_i - C_i(S_i(p_i))) + \sum_{m\in\mathcal{M}}(P^F_{fm} - p_m)Q^F_{fm}\right]$$

The terms within the expectation operator can be split into three distinct expectation terms:

$$\max_{\{S_i|i\in\mathcal{I}_f\}} \mathbb{E}\left[\sum_{i\in\mathcal{I}_f}(S_i(p_i)p_i - C_i(S_i(p_i)))\right] + \mathbb{E}\left[\sum_{m\in\mathcal{M}} P^F_{fm}Q^F_{fm}\right] - \mathbb{E}\left[\sum_{m\in\mathcal{M}} p_m Q^F_{fm}\right]$$

[3]See Cramton (2017) for a more detailed discussion of offer price mitigation in ERCOT.

The first term is the firm's expected profit from electricity production in the real-time wholesale market. The second term is the firm's revenue from existing forward positions. The third term is the firm's expected cost of effectively unwinding those forward positions in the real-time market. These three terms can be further rewritten as:

$$\max_{\{S_i|i\in\mathcal{I}_f\}} \sum_{i\in\mathcal{I}_f} \mathbb{E}\left[\pi_i\right] + \mathbb{E}_{\mathcal{M}}\left[\sum_{m\in\mathcal{M}} P^F_{fm}Q^F_{fm}\right] - \mathbb{E}_{\mathcal{M}}\left[\sum_{m\in\mathcal{M}} Q^F_{fm}\,\mathbb{E}_{-\mathcal{M}}\left[p_m \mid \mathcal{M}\right]\right]$$

In the first term, π_i is the real-time profit from production at unit i, as previously defined. Additionally, the order of the summation and expectation operators can be reversed due to the linearity of expectations. In the second term, the terms of forward positions are exogenous to the real-time market, so the only source of uncertainty is the market in which each position is taken; I denote that the expectation is taken only over the formation of local markets by the subscript $\mathcal{M}$ on the expectation operator. In the third term, I apply the Law of Total Expectation to separate uncertainty due to the formation of local markets from the remaining sources of uncertainty while conditioning on local markets; the forward position in each market depends only on local market formation, while the market-clearing prices depend on all sources of uncertainty.

The firm chooses its offer curves, $\{S_i \mid i \in \mathcal{I}_f\}$, to maximize this total profit. Each offer curve is defined by a vector of offer prices, $\mathbf{b}_i$, and a vector of offer quantities, $\mathbf{q}_i$. I assume the firm uses sufficiently many quantities and chooses an offer curve by selecting the optimal price at each quantity. Then the first-order condition with respect to the kth offer price at unit i is:

$$0 = \sum_{j\in\mathcal{I}_f} \frac{\partial\,\mathbb{E}\left[\pi_j\right]}{\partial b_{ik}} + \frac{\partial\,\mathbb{E}_{\mathcal{M}}\left[\sum_{m\in\mathcal{M}} P^F_{fm}Q^F_{fm}\right]}{\partial b_{ik}} - \frac{\partial\,\mathbb{E}_{\mathcal{M}}\left[\sum_{m\in\mathcal{M}} Q^F_{fm}\,\mathbb{E}_{-\mathcal{M}}\left[p_m \mid \mathcal{M}\right]\right]}{\partial b_{ik}}$$

In all three terms, the derivative can be moved inside the innermost expectation operator. Note, however, that the forward position prices and quantities in the second term do not vary with the real-time market, so the second term is equal to zero. Then the first-order condition reduces to:

$$0 = \sum_{j\in\mathcal{I}_f} \mathbb{E}\left[\frac{\partial \pi_j}{\partial b_{ik}}\right] - \mathbb{E}_{\mathcal{M}}\left[\sum_{m\in\mathcal{M}} Q^F_{fm}\,\mathbb{E}_{-\mathcal{M}}\left[\frac{\partial p_m}{\partial b_{ik}} \mid \mathcal{M}\right]\right] \tag{A.1}$$

I consider each of these two terms separately.

The first term in Equation (A.1) incorporates the change in expected profit at unit j due to a marginal change in the kth offer price at unit i. I assume no constraints are marginally binding, so a marginal change in b_{ik} will only affect π_j if unit j is in the same local market as unit i. Thus, the summation occurs over only the firm's plants within market m, which I use to denote unit i's local market. Additionally, a marginal change in b_{ik} will only affect other units if that offer price clears the markets, so I evaluate these terms at a price of b_{ik}. Finally, I expand π_j as defined previously:

$$\sum_{j\in\mathcal{I}_f} \mathbb{E}\left[\frac{\partial \pi_j}{\partial b_{ik}}\right] = \mathbb{E}\left[\sum_{j\in\mathcal{I}_f\cap\mathcal{I}_m} \frac{\partial}{\partial b_{ik}}\left(S_j(b_{ik})b_{ik} - C_j(S_j(b_{ik}))\right)\right]$$

Taking the derivative within the summation yields:

$$\sum_{j\in\mathcal{I}_f}\mathbb{E}\left[\frac{\partial\pi_j}{\partial b_{ik}}\right]=\mathbb{E}\left[\sum_{j\in\mathcal{I}_f\cap\mathcal{I}_m}\left(S_j(b_{ik})+\frac{\partial S_j(b_{ik})}{\partial b_{ik}}b_{ik}-C_j'(S_j(b_{ik}))\frac{\partial S_j(b_{ik})}{\partial b_{ik}}\right)\right]$$

A change in an offer price at unit i will change production only at unit i and not at producing units that are inframarginal, so the second and third terms are zero for all units other than unit i. I further combine these terms, which gives:

$$\sum_{j\in\mathcal{I}_f}\mathbb{E}\left[\frac{\partial\pi_j}{\partial b_{ik}}\right]=\mathbb{E}\left[\sum_{j\in\mathcal{I}_f\cap\mathcal{I}_m}S_j(b_{ik})+\frac{\partial S_i(b_{ik})}{\partial b_{ik}}\left(b_{ik}-C_i'(S_i(b_{ik}))\right)\right]$$

Every local market clears at the price that equates production and residual demand at every unit, as given in Equation (1.2). Thus, I replace production at unit i with residual demand at unit i. Additionally, I assume marginal cost is constant in quantity and given by c_i. Then this term can be rewritten as:

$$\sum_{j\in\mathcal{I}_f}\mathbb{E}\left[\frac{\partial\pi_j}{\partial b_{ik}}\right]=\mathbb{E}\left[\sum_{j\in\mathcal{I}_f\cap\mathcal{I}_m}S_j(b_{ik})\right]+(b_{ik}-c_i)\,\mathbb{E}\left[\frac{\partial R_i(b_{ik})}{\partial p_m}\right]\tag{A.2}$$

The second term in Equation (A.1) incorporates the change in the cost of unwinding forward positions due to a marginal change in the kth offer price at unit i. I again assume no constraints are marginally binding, so a marginal change in b_{ik} will only affect p_m, the price in unit i's local market. Conditional on the set of local markets, the quantity of the firm's forward position in unit i's local market is exogenous, but the firm has uncertainty over which local market will contain unit i. Thus, this term is the firm's expected forward position in unit i's local market, market m:

$$\mathbb{E}_{\mathcal{M}}\left[\sum_{m\in\mathcal{M}}Q_{fm}^F\,\mathbb{E}_{-\mathcal{M}}\left[\frac{\partial p_m}{\partial b_{ik}}\mid\mathcal{M}\right]\right]=\mathbb{E}\left[Q_{fm}^F\right]\tag{A.3}$$

Substituting Equations (A.2) and (A.3) into Equation (A.1) yields:

$$0=\mathbb{E}\left[\sum_{j\in\mathcal{I}_f\cap\mathcal{I}_m}S_j(b_{ik})\right]+(b_{ik}-c_i)\,\mathbb{E}\left[\frac{\partial R_i(b_{ik})}{\partial p_m}\right]-\mathbb{E}\left[Q_{fm}^F\right]$$

Combining the first and last terms, which are the quantities of electrcity sold in the real-time market and sold in advance through forward position, respectively, gives:

$$0=\mathbb{E}\left[\sum_{j\in\mathcal{I}_f\cap\mathcal{I}_m}S_j(b_{ik})-Q_{fm}^F\right]+(b_{ik}-c_i)\,\mathbb{E}\left[\frac{\partial R_i(b_{ik})}{\partial p_m}\right]$$

The first term is the net quantity of electricity sold by the firm in market m in the real-time market, which I previously defined as Q_{fm}. Substituting in $Q_{fm}(b_{ik})$ yields:

$$0 = \mathbb{E}\left[Q_{fm}(b_{ik})\right] + (b_{ik} - c_i)\,\mathbb{E}\left[\frac{\partial R_i(b_{ik})}{\partial p_m}\right]$$

I finally rewrite this expression to give the optimal markup rule, equivalent to Equation (1.3):

$$b_{ik} - c_i = \frac{\mathbb{E}[Q_{fm}(b_{ik})]}{-\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}$$

A.2.2 Comparative statics

In this section I derive the comparative static of the optimal markup rule, in Equation (1.3), with respect to the size of the local market in which the unit operates, K_i. Taking the derivative of the optimal markup rule with respect to the market size yields:

$$\frac{d(b_{ik} - c_i)}{dK_i} = \frac{-\frac{\partial\,\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\partial K_i} \times \mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right] + \frac{\partial\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}{\partial K_i} \times \mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]^2}$$

Splitting this result into two terms and cancelling common terms gives:

$$\frac{d(b_{ik} - c_i)}{dK_i} = \frac{-\frac{\partial\,\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\partial K_i}}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]} + \frac{\frac{\partial\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}{\partial K_i}}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]} \times \frac{\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]}$$

I then multiply both the numerator and denominator of the first term by the same expectation, which is the numerator of the optimal markup rule:

$$\frac{d(b_{ik} - c_i)}{dK_i} = \frac{-\frac{\partial\,\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\partial K_i}}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]} \times \frac{\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\mathbb{E}\left[Q_{fm}(b_{ik})\right]} + \frac{\frac{\partial\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}{\partial K_i}}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]} \times \frac{\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]}$$

Rearranging terms and negative signs gives:

$$\begin{aligned}\frac{d(b_{ik} - c_i)}{dK_i} = &\frac{\frac{\partial\,\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\partial K_i}}{\mathbb{E}\left[Q_{fm}(b_{ik})\right]} \times \frac{\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{-\,\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]} \\ &- \frac{\frac{\partial\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}{\partial K_i}}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]} \times \frac{\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{-\,\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]}\end{aligned}$$

Both terms now include the optimal markup rule, which I replace with the markup:

$$\frac{d(b_{ik} - c_i)}{dK_i} = \frac{\frac{\partial\,\mathbb{E}\left[Q_{fm}(b_{ik})\right]}{\partial K_i}}{\mathbb{E}\left[Q_{fm}(b_{ik})\right]}(b_{ik} - c_i) - \frac{\frac{\partial\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}{\partial K_i}}{\mathbb{E}\left[\partial R_i(b_{ik})/\partial p_m\right]}(b_{ik} - c_i)$$

Dividing by the markup yields the comparative static, equivalent to Equation (1.4):

$$\frac{d(b_{ik} - c_i)/dK_i}{b_{ik} - c_i} = \frac{\partial\,\mathbb{E}[Q_{fm}(b_{ik})]/\partial K_i}{\mathbb{E}[Q_{fm}(b_{ik})]} - \frac{\partial\,\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]/\partial K_i}{\mathbb{E}[\partial R_i(b_{ik})/\partial p_m]}$$

A.3 Data

A.3.1 Marginal cost

I first calculate heat rate for natural gas and coal units. From the EPA's CEMS data, I observe hourly fuel input, which is measured as the heat content of the fuel being consumed at the unit in that hour. These data are reported only for units larger than 25 MW. From ERCOT's 60-Day SCED Disclosure Reports, I observe net generation for all units at 15-minute intervals. From these two data sources, I construct an hourly measure of heat rate — the ratio of fuel input to net generation, which is an inverse measure of unit efficiency — at units larger than 25 MW. These data, however, exhibit surprising hourly variation, which is not to be expected; a unit is typically expected to operate within a relatively small range of heat rates based on its design characteristics. It is not clear if this heat rate variation is real or due to measurement error, which can be amplified when taking a ratio. Even if this variation is real, it is not clear that firms would expect this hourly variation in heat rates and incorporate it into the unit's offer curves. Thus, I use these hourly heat rate data only as a robustness check. Instead, for each unit larger than 25 MW, I aggregate these high-frequency data to calculate the unit's average heat rate over all four years of my data. I use this average heat rate when calculating the unit's marginal cost in my main analysis.

This procedure, however, produces average heat rates only for units larger than 25 MW. For smaller units, I use the EIA's Form 923, which reports monthly fuel input and net generation at each plant[4] by prime mover and fuel type. As with the larger units, I aggregate these data over all four years and then take the ratio of fuel input to net generation to calculate the average heat rate. I apply this heat rate, averaged over a plant, prime mover, and fuel type, to all applicable units that do not have sufficient data to calculate heat rate using the prior method. Several units, however, have insufficient data using both of these methods. So I finally use the Form 923 data to calculate the average heat rate for every prime mover and fuel type combination based on the plants that do have data available. For the units for which I cannot otherwise calculate a marginal cost, I apply the market-wide average heat rate for the unit's prime mover and fuel type.

I next calculate emissions rates for natural gas and coal units using a procedure similar to that for heat rates. The EPA's CEMS data report hourly emissions of SO_2 and NO_x. As with heat rate, I use these data and ERCOT's net generation data to calculate the average emissions rates over all four years of data for units larger than 25 MW. In my data, all units smaller than 25 MW use natural gas as the primary fuel. I approximate the average

[4]These data sources highlight an important difference between a "unit" and a "plant." A plant may comprise two or more units. For example, if one facility houses four natural gas turbines, each of those turbines is a unit, while the entire facility is a plant. The EIA's Form 923 reports data at this more aggregated level of the plant with the data split out by different prime movers — broad classifications of generation technologies — and fuel type. It is important to note that in the main text of this paper, I use the terms "power plant" and "unit" interchangeably to describe the cross-sectional unit of my analysis, which I keep as disaggregated as possible while still allowing me to accurately join different data sources.

emissions rates at these units as the product of the pollutant content of natural gas[5] and the average heat rate at the unit, as previously calculated.

I next calculate the fuel cost of natural gas and coal units using their average heat rates and fuel prices. For natural gas units, I observe daily natural gas prices, from S&P Global Market Intelligence, at many hubs throughout the state of Texas. I follow Hortaçsu and Puller (2008) and select one natural gas hub for each ERCOT load zone[6] and apply its price to all natural gas units located in that load zone. I also add a distribution charge of $0.10 per MMBtu of natural gas. For coal units, I calculate the monthly average spot price of coal using the EIA's Form 923, which reports the quantity and total cost of coal shipments to power plants. I restrict these data to spot purchases of coal by plants in ERCOT and then calculate the average price for each month in my data. I apply that monthly average coal price to all coal units for that month. To calculate the fuel cost at these units, I take the product of the unit's heat rate and fuel price.

I next calculate the emissions cost of natural gas and coal units using their average emissions rates and permit prices. The EPA's Air Market Program Data (AMPD) report which units are subject to emissions regulations and which pollutants are subject to regulation at those units. I calculate the emissions cost only for pollutants that are regulated at a unit. From S&P Global Market Intelligence, I observe daily SO_2 permit prices for all four years in my data and daily NO_x permit prices beginning in December 2012. I supplement these daily prices with monthly NO_x permit prices from the EPA's market progress reports.[7] I calculate the emissions cost for each applicable pollutant as the product of a unit's emissions rate and permit price, and then sum over all applicable pollutants.

I finally calculate the marginal cost at every unit. For natural gas and coal units, I sum the fuel costs and emissions costs that I previously calculated. For nuclear units, I use estimates of nuclear operating costs from Deutch et al. (2009). For all other units, which are almost entirely wind units, I assume marginal cost is zero.

This marginal cost calculation is relatively simple and straightforward, especially when compared to estimation procedures that are required in other settings. Concern may exist, however, if these components of marginal cost are mis-measured or if additional components of marginal cost are omitted; for example, all units have a small variable cost of operations and maintenance, which should be included in this marginal cost. My empirical strategy addresses these concerns through fixed effects. A time-invariant error in a unit's marginal cost is accounted for in a unit fixed effect, while an error common to all units is accounted for in an hour of sample fixed effect. Additionally, to check for robustness of my results, I remove marginal cost from my analysis and estimate the effect of local market size on offer prices, rather than on markups.

[5] I assume 0.0006 lb of SO_2 and 0.17 lb of NO_x per Mcf of natural gas, from the EPA's eGRID technical support documentation, available at `https://www.epa.gov/sites/production/files/2018-02/documents/egrid2016_technicalsupportdocument.pdf`.

[6] I use Waha for the West load zone, Carthage for the North load zone, Agua Dulce for the South load zone, and Katy for the Houston load zone

[7] Available at `www.epa.gov/airmarkets/nox-budget-trading-program-historical-reports`.

A.3.2 Temperature

I use temperature data from the National Oceanic and Atmospheric Administration's (NOAA) Integrated Surface Database. I restrict my dataset to weather stations in the state of Texas that have temperature readings for at least 97.5% of the hours during my study period and no more than 60 consecutive hours of missing data, which reduces the extent of missing data. Some of these weather stations record temperature more frequently than once an hour, in which case I average temperatures to the hourly level. The resulting dataset is missing less than 0.5% of hourly temperature data.

I impute these missing temperature data to create a balanced panel of hourly temperatures at 75 weather stations in the state of Texas. For each missing temperature, I fit a model of that weather station's temperature in that particular hour of day and month of year. I model the temperature as a function of temperatures at the same weather station in the closest preceding and successive hours that have non-missing data and of contemporaneous temperatures at the three weather stations that exhibit the highest temperature correlation with the station where temperature is being imputed. That is, using non-missing observations in the same hour of day and month of year as the temperature to be imputed, I estimate the model:

$$T_{it} = \beta_0 + \beta_1 T_{it'} + \beta_2 T_{it'}^2 + \beta_3 T_{it''} + \beta_4 T_{it''}^2 + \beta_5 T_{jt} + \beta_6 T_{jt}^2 + \beta_7 T_{j't} + \beta_8 T_{j't}^2 + \beta_9 T_{j''t} + \beta_{10} T_{j''t}^2$$

where t' is the closest preceding hour with non-missing data, t'' is the closest successive hour with non-missing data, j is the weather station that exhibits the greatest temperature correlation with weather station i in this hour of day and month of year, and j' and j'' are the two next most correlative weather stations.

For example, consider a weather station missing temperature data for the 3 and 4 p.m. hours on a particular day in July. First, using observations that are not missing, I estimate a model of the weather station's temperatures during the 3 p.m. hour in July as a function of that station's temperature in the 2 p.m. and 5 p.m. hours[8] on the same day and the contemporaneous temperatures at the three stations that exhibit the greatest temperature correlation with this station in the 3 p.m. hour in July. I then use the estimated parameters of that model and the comparable temperatures for the hour to be imputed in order to predict the missing temperature. I also perform a similar procedure to impute the missing temperature in the 4 p.m. hour. Note that this model will be different because it uses non-missing temperatures during the 4 p.m. hour in July and the most correlative weather stations may differ for the 4 p.m. hour in July.

A.4 Results

A.4.1 Decomposition of effects

Table 1.6 in the main text shows the effect of local market capacity on markups at 85% capacity decomposed by owner and unit type. Tables A.1 and A.2 display analogous results

[8] I do not use the 4 p.m. hour because that temperature is also missing, so I cannot use it to impute the 3 p.m. hour.

for the number of units in the local market and the number of firms in the local market, respectively. The outcome of this decomposition is consistent across all three measures of local market size: the main results are driven by competing firms and dispatchable units, capacity or units owned by the same firm have an opposite and larger effect, and non-dispatchable units have no effect. Tables A.3–A.5 show similar results when using temperature bins in the first stage or when examining markups at 75% capacity or 65% capacity. Figures A.1–A.3 display the first stages for these regressions, or the temperature-response functions for these measures of local market size by owner or unit type.

Table A.1: Units in local market by owner and type — Effect on markups

	Markup at 85% capacity ($/MWh)		
	(1)	(2)	(3)
Units in local market	−10.27**		
	(5.06)		
Units owned by competing firms (GW)		−14.44**	
		(6.40)	
Units owned by own firm (GW)		67.78*	
		(35.01)	
Dispatchable units			−13.16**
			(5.72)
Non-dispatchable units			1.66
			(9.65)
First-stage specification	Spline	Spline	Spline
First-stage F-statistic	4.98	5.36, 5.92	6.43, 6.50
Local temperature	X	X	X
Unit × hour × month FEs	X	X	X
Hour of sample FEs	X	X	X
Observations	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with measures of the number of units in the local market as the independent variables. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.4. Column (2) reports results with the number of units in the local market split by ownership. Column (3) reports results with the number of units in the local market split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for Columns (2) and (3) are displayed in Figure A.2. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Firms in local market by unit type — Effect on markups

	Markup at 85% capacity ($/MWh)	
	(1)	(2)
Firms in local market	−41.10**	
	(19.17)	
Firms with dispatchable units		−48.21**
		(20.20)
Firms with non-dispatchable units		2.50
		(32.73)
First-stage specification	Spline	Spline
First-stage F-statistic	5.44	7.03, 5.54
Local temperature	X	X
Unit × hour × month FEs	X	X
Hour of sample FEs	X	X
Observations	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with measures of the number of firms in the local market as the independent variables. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.5. Column (2) reports results with the number of firms in the local market split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for Column (2) are displayed in Figure A.3. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.3: Local market capacity by owner and unit type — Effect on markups using temperature bins or alternate quantities

	Markup at 85% capacity ($/MWh)			Markup at 75% capacity ($/MWh)			Markup at 65% capacity ($/MWh)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Local market capacity (GW)	−18.35**			−15.24**			−13.00***		
	(7.63)			(6.46)			(5.03)		
Capacity owned by competing firms (GW)		−45.77**			−42.33			−30.25	
		(19.88)			(28.66)			(19.66)	
Capacity owned by own firm (GW)		455.43**			524.11			341.91	
		(205.92)			(401.75)			(299.37)	
Dispatchable capacity (GW)			−28.89***			−19.17**			−12.28**
			(11.13)			(9.11)			(5.84)
Non-dispatchable capacity (GW)			65.68			29.99			−21.63
			(42.74)			(64.38)			(43.34)
First-stage specification	Bins	Bins	Bins	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	2.93	2.88, 2.97	2.98, 2.51	5.88	6.06, 3.66	6.36, 5.04	5.28	5.46, 3.46	5.69, 4.82
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating regressions similar to Equation (1.10) with measures of local market capacity as the independent variables. The dependent variable is markup at three different quantities of production, as shown in the column headers. Columns (1), (4), and (7) duplicate Columns (2), (6), and (9), respectively, of Table 1.3. Columns (2), (5), and (8) report results with local market capacity split by ownership. Columns (3), (6), and (9) report results with local market capacity split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for the split measures are displayed in Figure A.1. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4: Units in local market by owner and type — Effect on markups using temperature bins or alternate quantities

	Markup at 85% capacity ($/MWh)			Markup at 75% capacity ($/MWh)			Markup at 65% capacity ($/MWh)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Units in local market	−5.74**			−5.40**			−5.30**		
	(2.79)			(2.58)			(2.21)		
Units owned by competing firms		−8.48**			−10.36*			−9.43**	
		(3.70)			(5.58)			(4.27)	
Units owned by own firm		49.50*			85.03			70.83	
		(26.54)			(67.72)			(60.22)	
Dispatchable units			−11.54**			−7.93**			−6.24**
			(4.72)			(3.38)			(2.46)
Non-dispatchable units			10.55*			5.22			−1.76
			(6.27)			(8.44)			(5.59)
First-stage specification	Bins	Bins	Bins	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	2.11	2.27, 2.22	2.61, 3.02	4.65	5.29, 5.39	5.90, 6.68	4.29	4.91, 5.19	5.20, 6.63
Local temperature	X	X	X	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	5,224,152	5,224,152	5,224,152	5,415,331	5,415,331	5,415,331

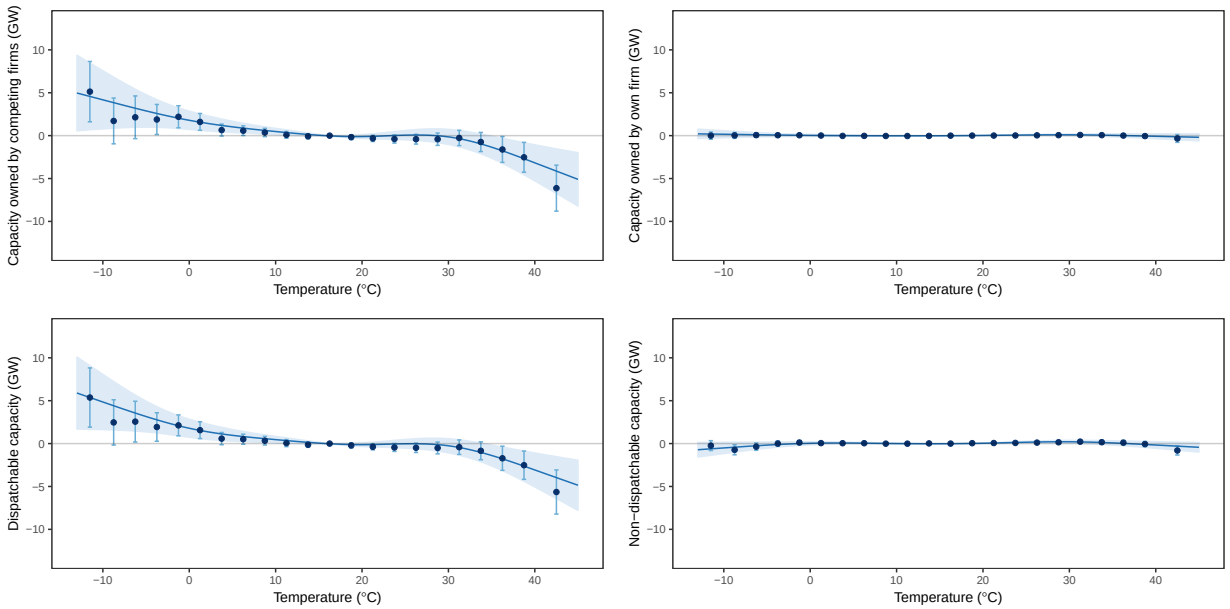
Notes: This table reports results of estimating regressions similar to Equation (1.10) with measures of the number of units in the local market as the independent variables. The dependent variable is markup at three different quantities of production, as shown in the column headers. Columns (1), (4), and (7) duplicate Columns (2), (6), and (9), respectively, of Table 1.4. Columns (2), (5), and (8) report results with the number of units in the local market split by ownership. Columns (3), (6), and (9) report results with the number of units in the local market split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for the split measures are displayed in Figure A.2. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Firms in local market by unit type — Effect on markups using temperature bins or alternate quantities

	Markup at 85% capacity ($/MWh)		Markup at 75% capacity ($/MWh)		Markup at 65% capacity ($/MWh)	
	(1)	(2)	(3)	(4)	(5)	(6)
Firms in local market	−23.82**		−22.95**		−21.64**	
	(10.64)		(10.25)		(8.47)	
Firms with dispatchable units		−44.85**		−31.90**		−23.34**
		(17.44)		(14.27)		(9.40)
Firms with non-dispatchable units		31.47*		18.18		−5.97
		(18.65)		(28.93)		(19.00)
First-stage specification	Bins	Bins	Spline	Spline	Spline	Spline
First-stage F-statistic	2.63	3.22, 2.92	5.20	6.55, 5.98	4.87	5.93, 5.70
Local temperature	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	4,820,775	4,820,775	5,224,152	5,224,152	5,415,331	5,415,331

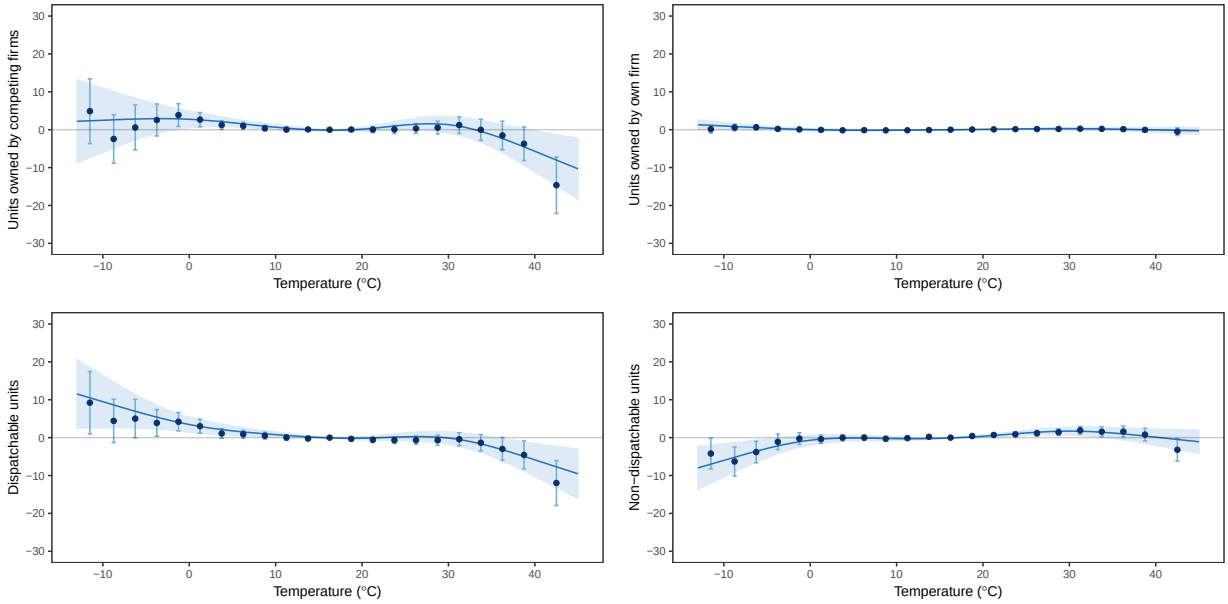
Notes: This table reports results of estimating regressions similar to Equation (1.10) with measures of the number of firms in the local market as the independent variables. The dependent variable is markup at three different quantities of production, as shown in the column headers. Columns (1), (3), and (5) duplicate Columns (2), (6), and (9), respectively, of Table 1.5. Columns (2), (4), and (6) report results with the number of firms in the local market split by unit type. All regressions are estimated with 2SLS and use a natural cubic spline in the first stages; first-stage results for the split measures are displayed in Figure A.3. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A.1: Temperature-response functions for local market capacity by owner or unit type



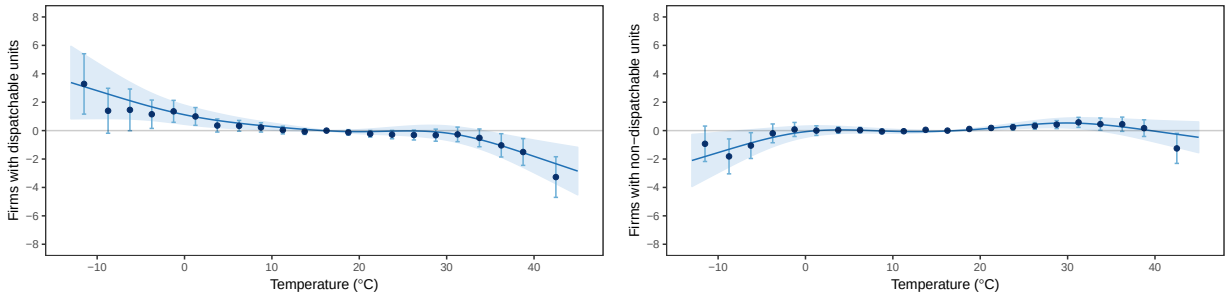
Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (1.11) for a different measure of local market capacity, as denoted on the y-axis of each panel. The top-left panel plots local market capacity owned by competing firms, the top-right panel plots local market capacity owned by the plant's own firm, the bottom-left panel plots local market capacity that is dispatchable, and the bottom-right panel plots local market capacity that is non-dispatchable. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.2: Temperature-response functions for units in local market by owner or unit type



Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (1.11) for a different measure of the number of units in the local market, as denoted on the y-axis of each panel. The top-left panel plots the number of units in the local market owned by competing firms, the top-right panel plots the number of units in the local market owned by the plant's own firm, the bottom-left panel plots the number of units in the local market that are dispatchable, and the bottom-right panel plots the number of units in the local market that are non-dispatchable. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.3: Temperature-response functions for firms in local market by unit type



Notes: Each panel of this figure plots a temperature-response function that results from estimating Equation (1.11) for a different measure of the number of firms in the local market, as denoted on the y-axis of each panel. The left panel plots the number of firms in the local market that own dispatchable units, and the right panel plots the number of firms in the local market that own non-dispatchable units. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

A.4.2 Robustness checks for local controls and temperature matching

Tables 1.7 and 1.8 in the main text show that the effect of local market capacity on markups at 85% capacity is robust to the inclusion of local market controls or to the use of a minimum matching distance. Tables A.6 and A.7 display analogous results for the number of units in the local market and the number of firms in the local market, respectively, which are also robust to these alternate specifications. Figure A.4 plots the estimated effects of local market size on markups when using temperature bins in the first stage or when examining markups at 75% capacity or 65% capacity. None of these results are statistically different from the comparable result presented in the main text; if anything, effects tend become larger when imposing a minimum matching distance.

Figures A.5–A.10 display the first stages for these regressions with alternate specifications. Figures A.5 and A.6 plot the temperature-response functions for local market capacity with other local controls and minimum matching distances, respectively. Figures A.7 and A.8 plot the temperature-response functions for the number of units in the local market with other local controls and minimum matching distances, respectively. Figures A.9 and A.10 plot the temperature-response functions for the number of firms in the local market with other local controls and minimum matching distances, respectively.

Figures A.11–A.16 display the reduced form for these regressions with alternate specifications. Figures A.11 and A.12 plot the temperature-response functions for markup at 85% capacity with other local controls and minimum matching distances, respectively. Figures A.13 and A.14 plot the temperature-response functions for markup at 75% capacity with other local controls and minimum matching distances, respectively. Figures A.15 and A.16 plot the temperature-response functions for markup at 65% capacity with other local controls and minimum matching distances, respectively.

Table A.6: Robustness checks — Effect of units in local market on markups

	Markup at 85% capacity ($/MWh)						
		Local controls			Minimum matching distance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Units in local market	−10.27**	−9.23**	−11.73**	−12.10*	−14.43**	−13.93**	−8.28**
	(5.06)	(3.90)	(5.18)	(7.13)	(6.24)	(5.83)	(3.61)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	4.98	6.94	7.57	5.30	5.72	7.20	5.86
Minimum matching distance (km)	0	0	0	0	100	200	300
Local temperature	X			X	X	X	X
Local capacity factor			X	X			
Unit × hour × month FEs	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

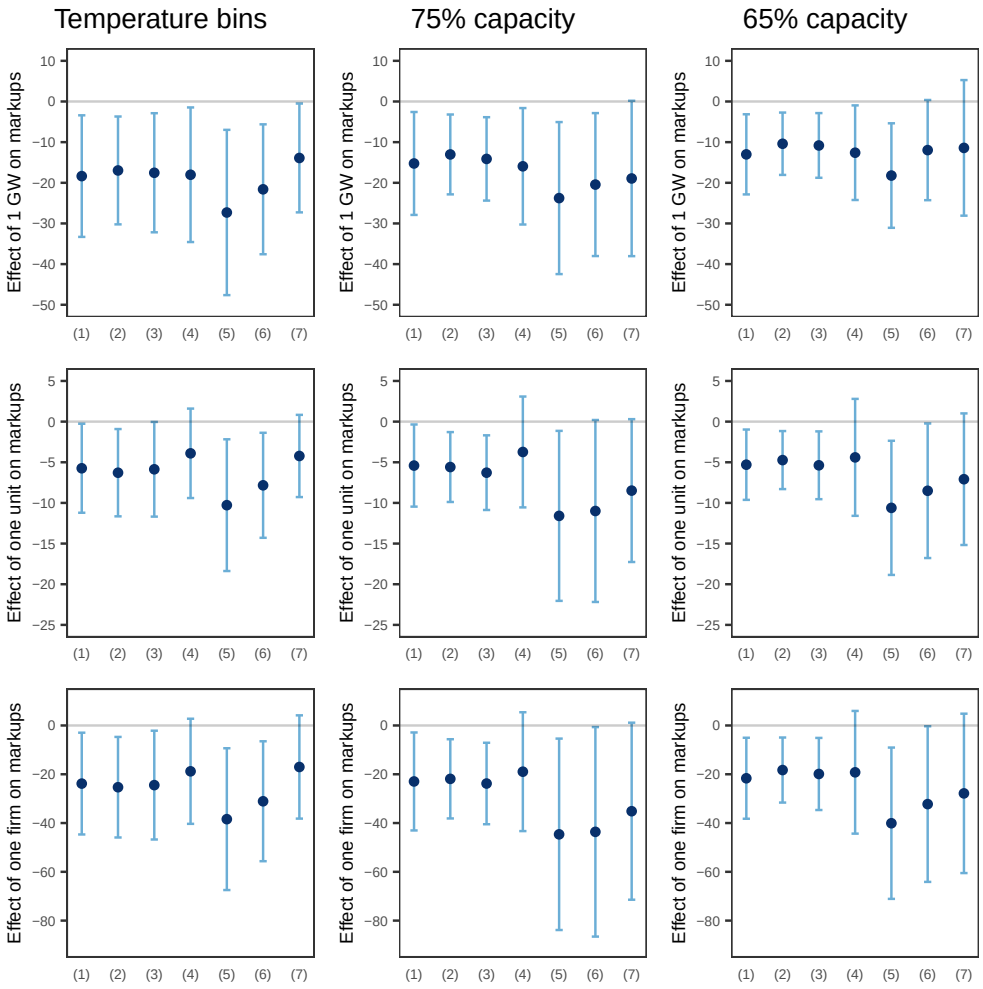
Notes: This table reports results of estimating regressions similar to Equation (1.10) with the number of units in the local market as the independent variable. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.4. Columns (2)–(4) report results with different sets of local market controls, as denoted in the table. Columns (5)–(7) report results with different matched weather stations; each unit is matched to the weather station that best predicts the unit's local market capacity and meets the minimum matching distance, as denoted in the table. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (2)–(7) are displayed in Figures A.7 and A.8. All regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.7: Robustness checks — Effect of firms in local market on markups

	Markup at 85% capacity ($/MWh)						
		Local controls			Minimum matching distance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Firms in local market	−41.10**	−35.53**	−43.27**	−49.18*	−56.10**	−56.32**	−34.43**
	(19.17)	(14.64)	(18.40)	(26.16)	(23.38)	(22.15)	(15.21)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	5.44	7.46	8.13	5.61	6.10	8.06	6.25
Minimum matching distance (km)	0	0	0	0	100	200	300
Local temperature	X			X	X	X	X
Local capacity factor			X	X			
Unit × hour × month FEs	X	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

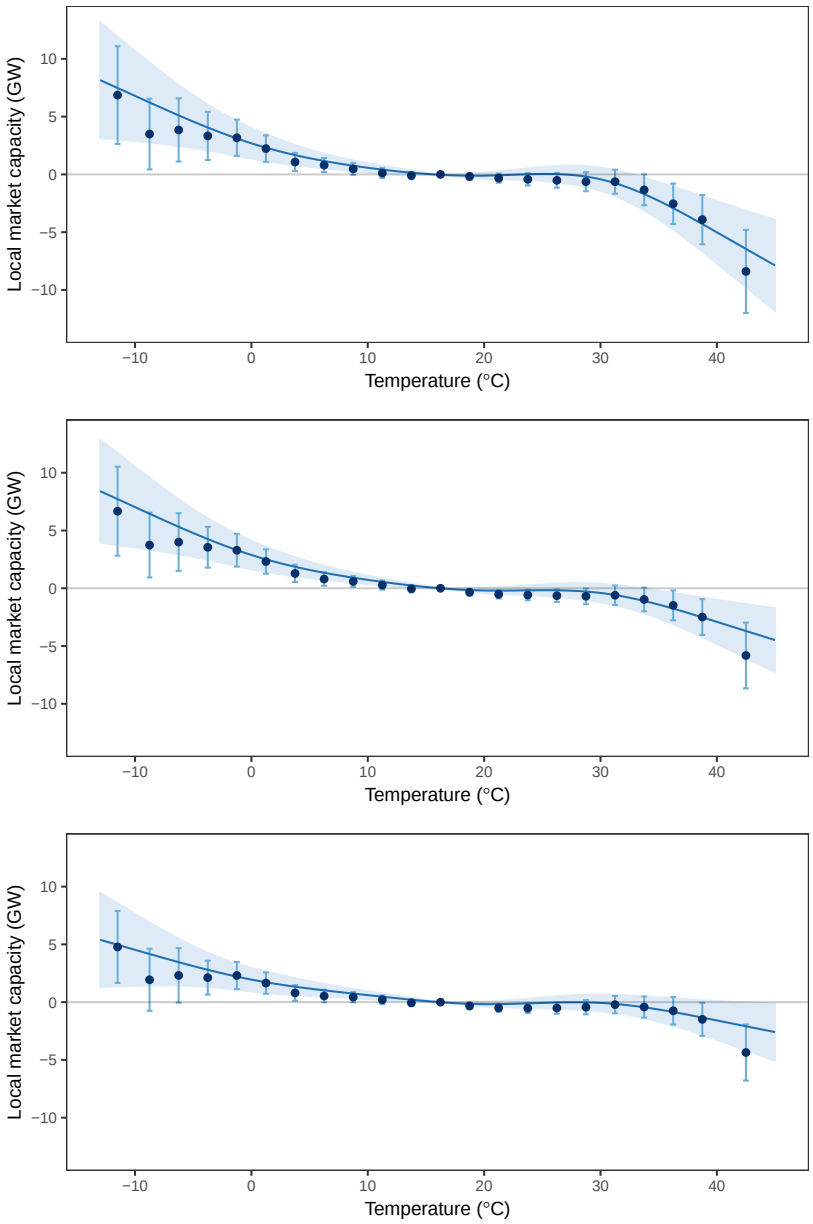
Notes: This table reports results of estimating regressions similar to Equation (1.10) with the number of firms in the local market as the independent variable. The dependent variable is markup at 85% capacity. Column (1) duplicates Column (3) of Table 1.5. Columns (2)–(4) report results with different sets of local market controls, as denoted in the table. Columns (5)–(7) report results with different matched weather stations; each unit is matched to the weather station that best predicts the unit's local market capacity and meets the minimum matching distance, as denoted in the table. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (2)–(7) are displayed in Figures A.9 and A.10. All regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A.4: Robustness checks — Effects on markups using temperature bins or alternate quantities



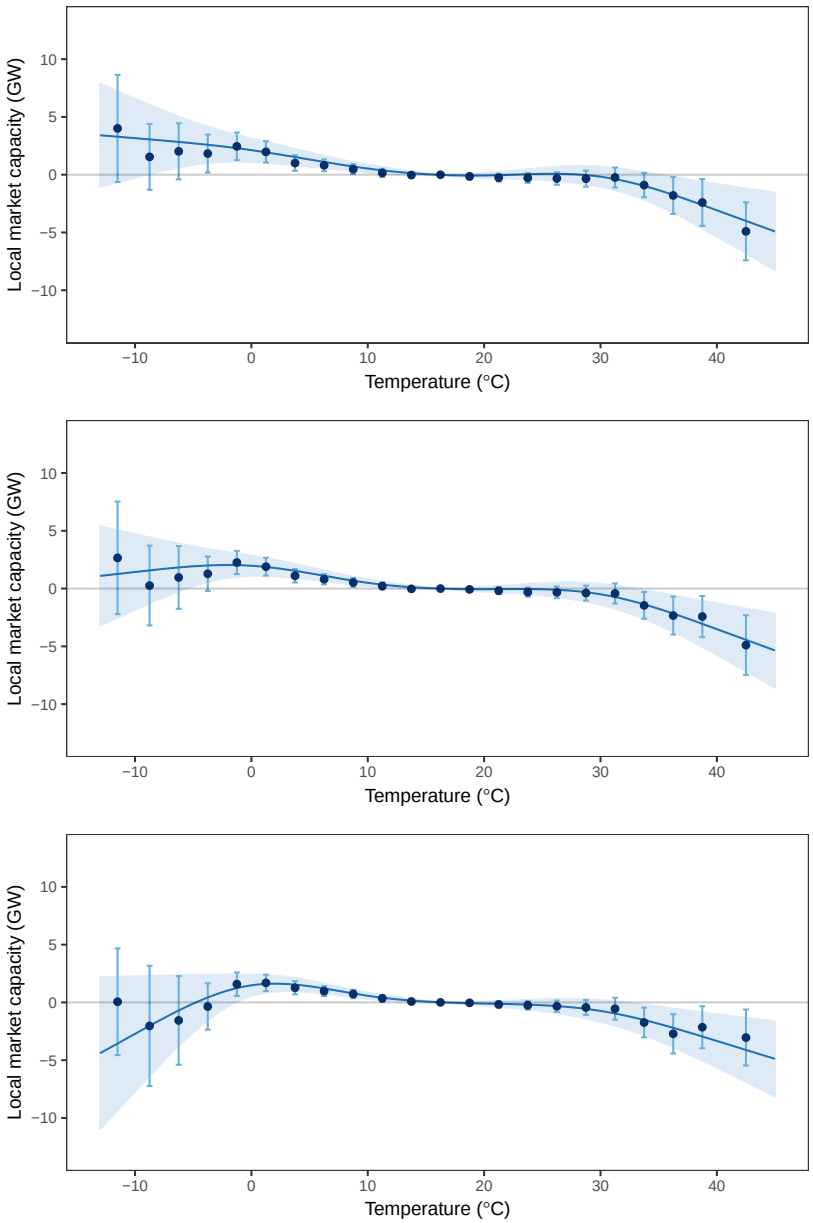
Notes: This figures displays results of estimating regressions similar to Equation (1.10). Each panel shows results from a set of regressions with the same independent variable and dependent variable; the independent variable varies by row and the dependent variable varies by column. From top to bottom, the independent variables are local market capacity, number of units in the local market, and number of firms in the local market. From left to right, the dependent variable are markups at 85% capacity, 75% capacity, and 65% capacity. Each panel displays the results of seven regressions, which are analogous to the columns of Tables A.6 and A.7: (1) duplicates the main result presented in Tables 1.3–1.5; (2) removes the control for local temperature, (3) replaces the control for local temperature with a control for local capacity factor, (4) adds a control for local capacity factor, (5) imposes a minimum matching distance of 100 km, (6) imposes a minimum matching distance of 200 km, and (7) imposes a minimum matching distance of 300 km. The points and error bars show the coefficient estimates and 95% confidence intervals from these regressions. All regressions are estimated with 2SLS; regressions for markups at 85% capacity use temperature bins in the first stage, and regressions for markups at 75% capacity and 65% capacity use a natural cubic spline in the first stage. All regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Confidence intervals use standard errors that are two-way clustered at the unit and date.

Figure A.5: Temperature-response functions for local market capacity with other local controls



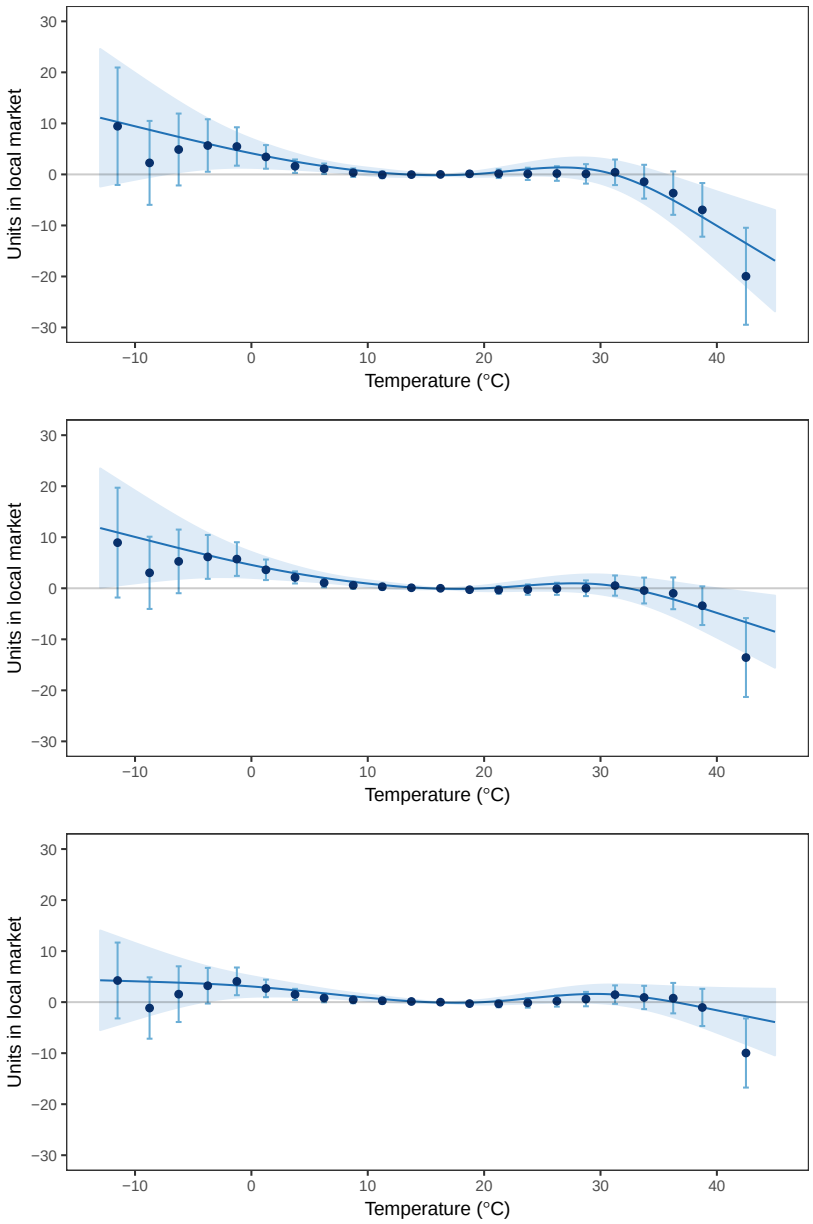
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.11) for local market capacity. Each panel displays the results from a different regression specification; from top to bottom, the regression removes the control for local temperature, replaces the control for local temperature with a control for local capacity factor, and adds a control for local capacity factor, as compared to Equation (1.11). The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.6: Temperature-response functions for local market capacity with minimum distance



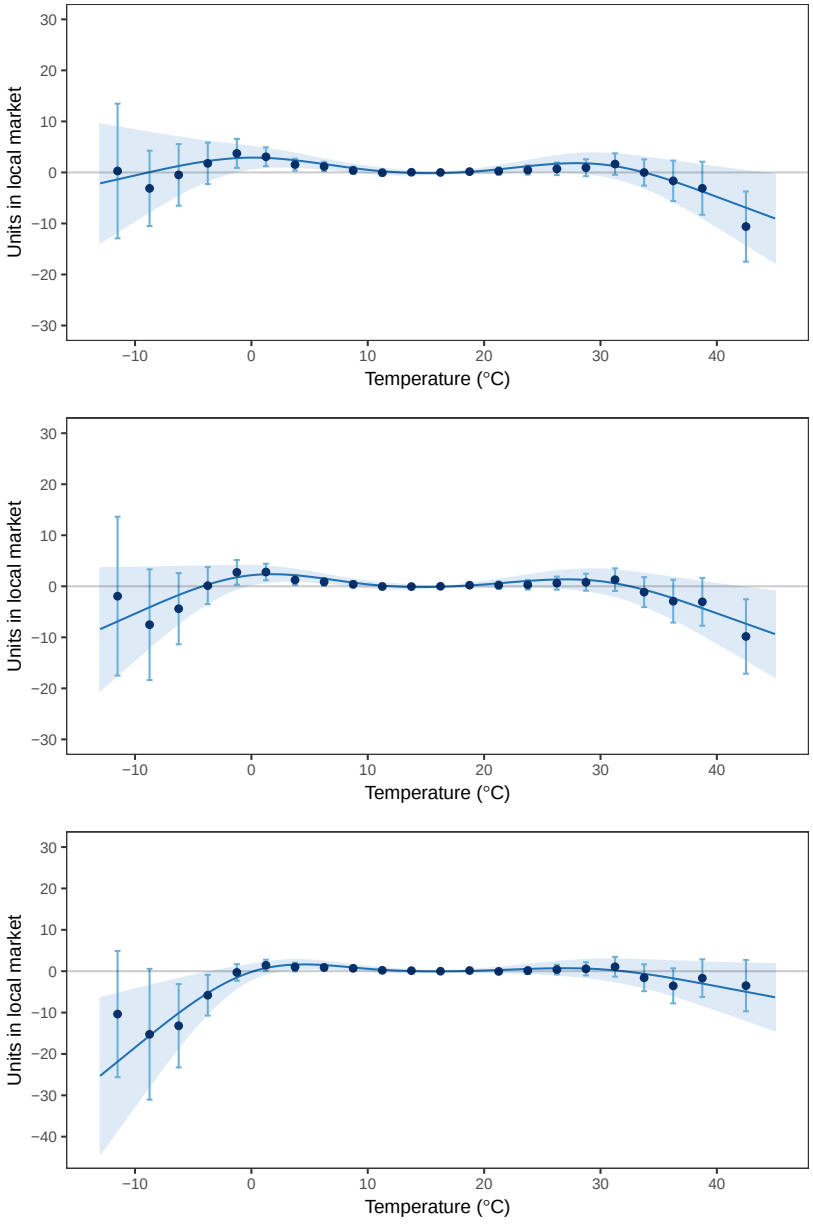
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.11) for local market capacity. Each panel displays the results from imposing a different minimum distance between each unit and its matched weather station; from top to bottom, the distances are 100 km, 200 km, and 300 km. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.7: Temperature-response functions for units in local market with other local controls



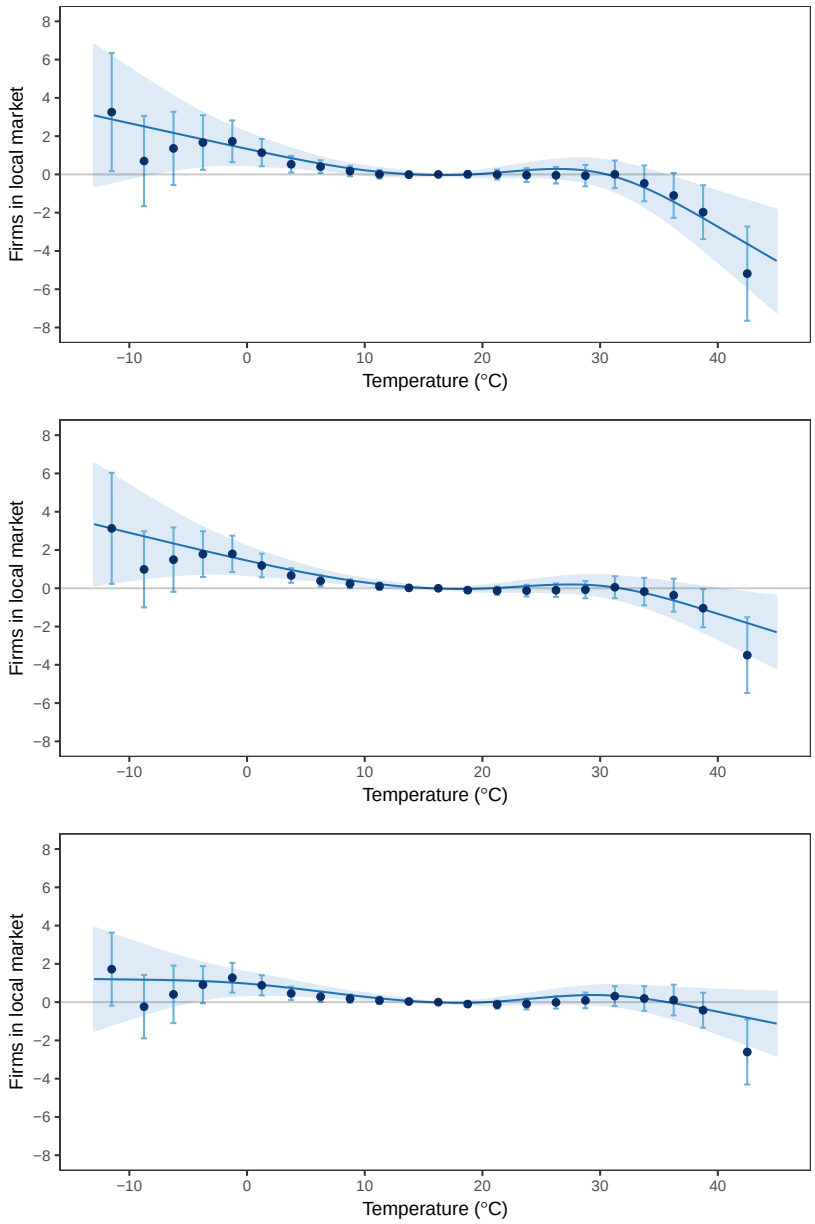
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.11) for the number of units in the local market. Each panel displays the results from a different regression specification; from top to bottom, the regression removes the control for local temperature, replaces the control for local temperature with a control for local capacity factor, and adds a control for local capacity factor, as compared to Equation (1.11). The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.8: Temperature-response functions for units in local market with minimum distance



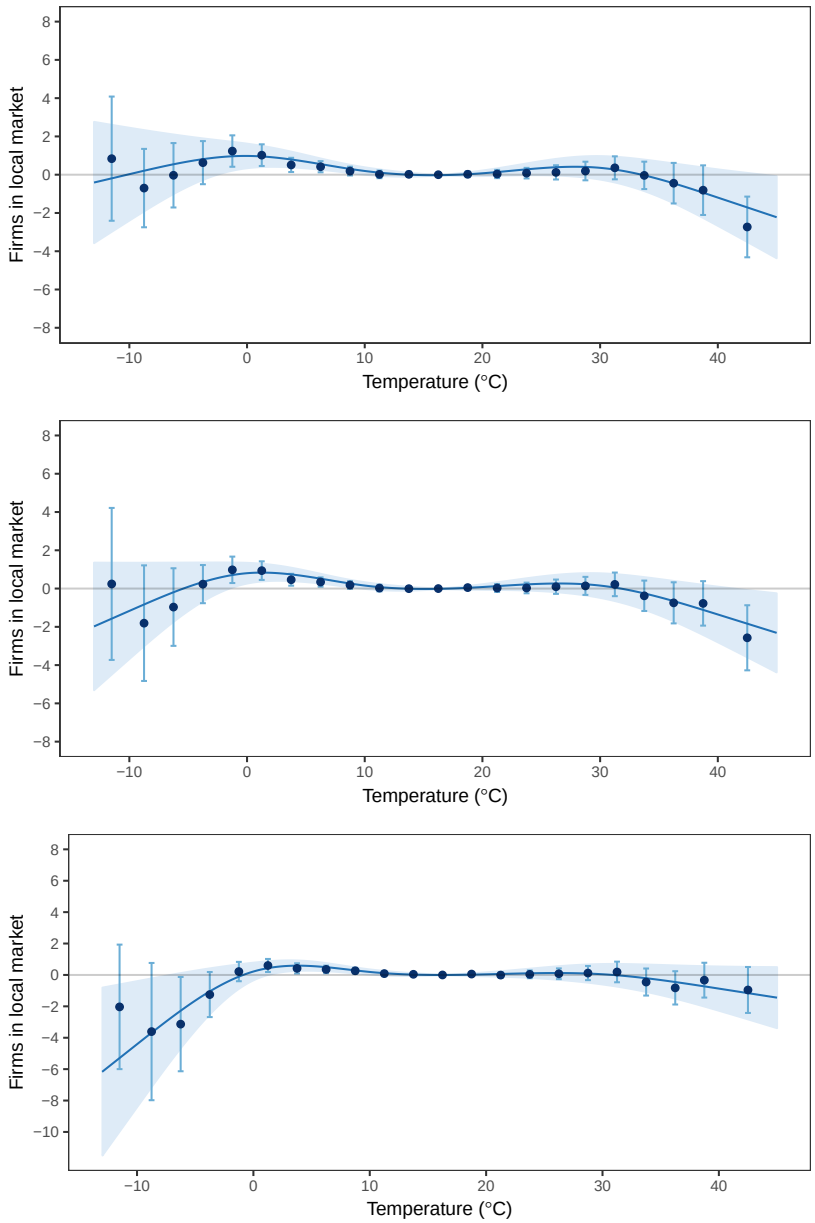
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.11) for the number of units in the local market. Each panel displays the results from imposing a different minimum distance between each unit and its matched weather station; from top to bottom, the distances are 100 km, 200 km, and 300 km. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.9: Temperature-response functions for firms in local market with other local controls



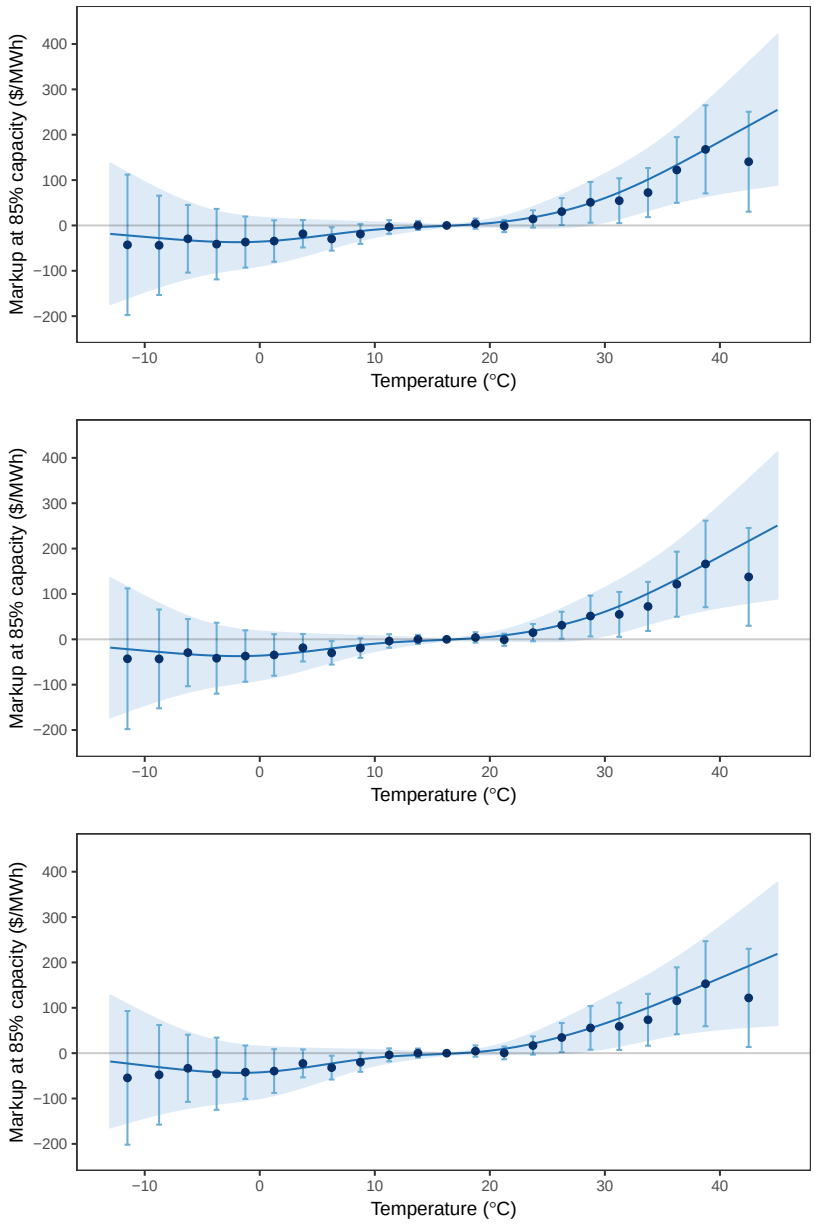
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.11) for the number of firms in the local market. Each panel displays the results from a different regression specification; from top to bottom, the regression removes the control for local temperature, replaces the control for local temperature with a control for local capacity factor, and adds a control for local capacity factor, as compared to Equation (1.11). The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.10: Temperature-response functions for firms in local market with minimum distance



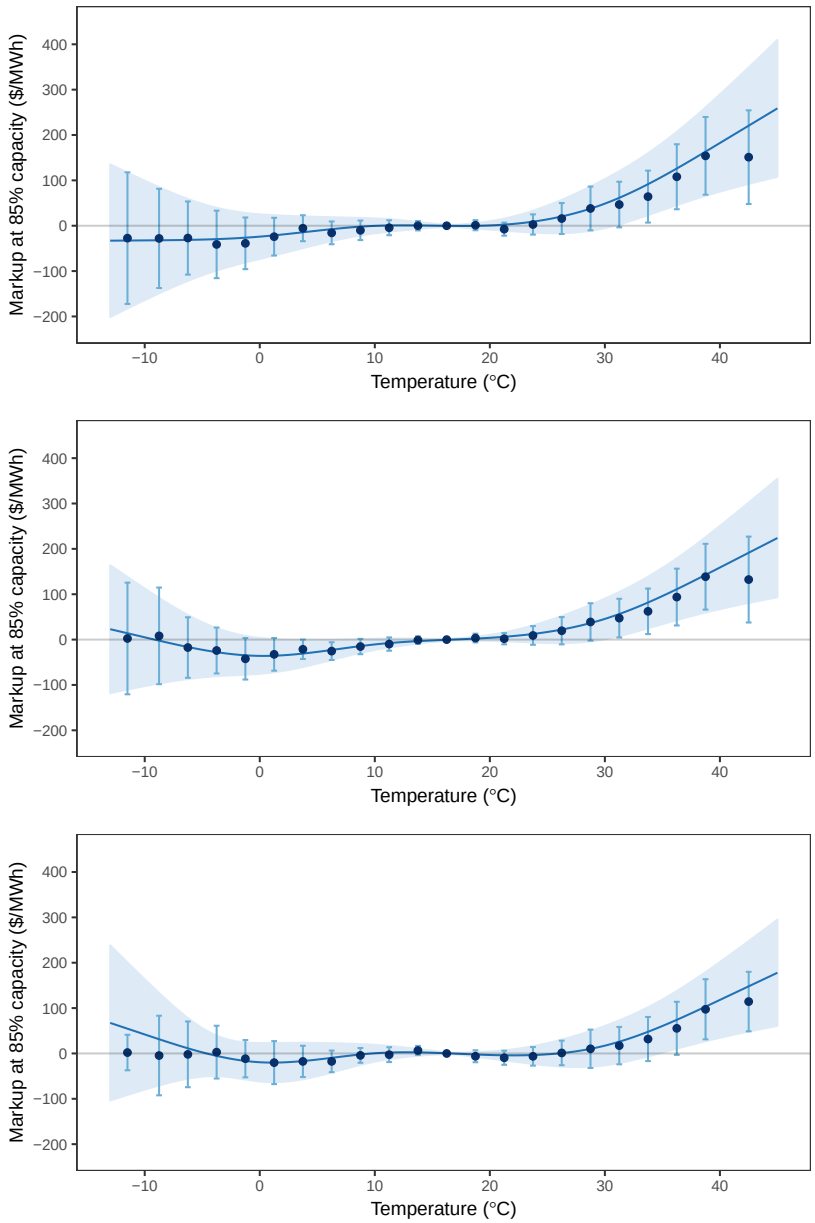
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.11) for the number of firms in the local market. Each panel displays the results from imposing a different minimum distance between each unit and its matched weather station; from top to bottom, the distances are 100 km, 200 km, and 300 km. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.11: Temperature-response functions for markup at 85% capacity with other local controls



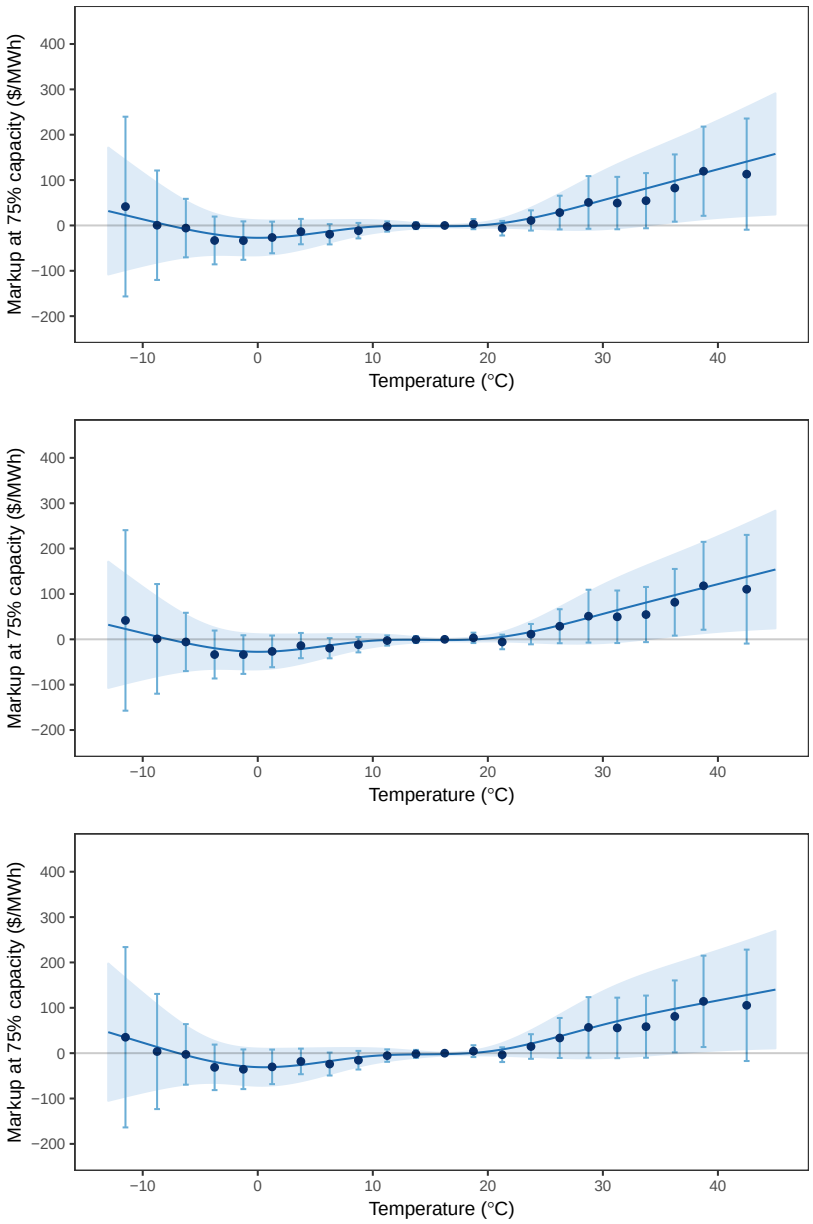
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for markup at 85% capacity. Each panel displays the results from a different regression specification; from top to bottom, the regression removes the control for local temperature, replaces the control for local temperature with a control for local capacity factor, and adds a control for local capacity factor, as compared to Equation (1.14). The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects.

Figure A.12: Temperature-response functions for markup at 85% capacity with minimum distance



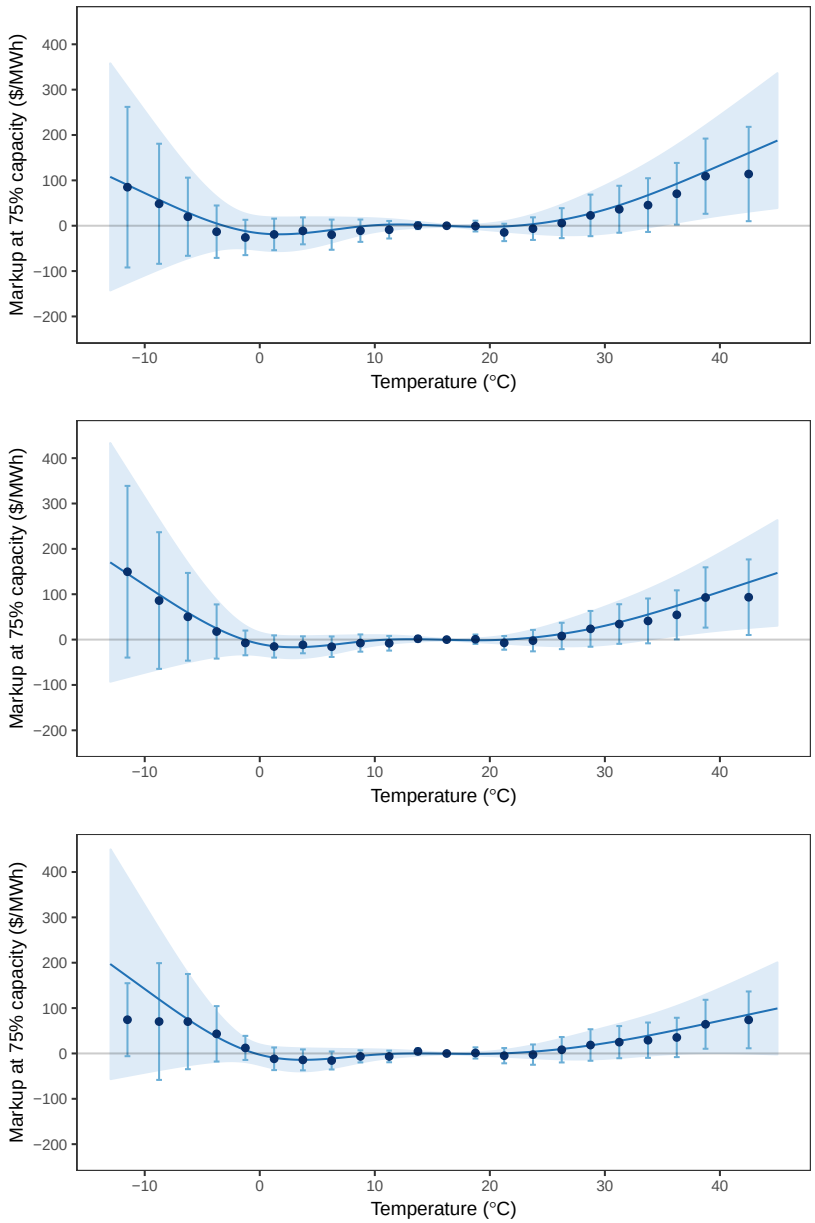
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for markup at 85% capacity. Each panel displays the results from imposing a different minimum distance between each unit and its matched weather station; from top to bottom, the distances are 100 km, 200 km, and 300 km. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.13: Temperature-response functions for markup at 75% capacity with other local controls



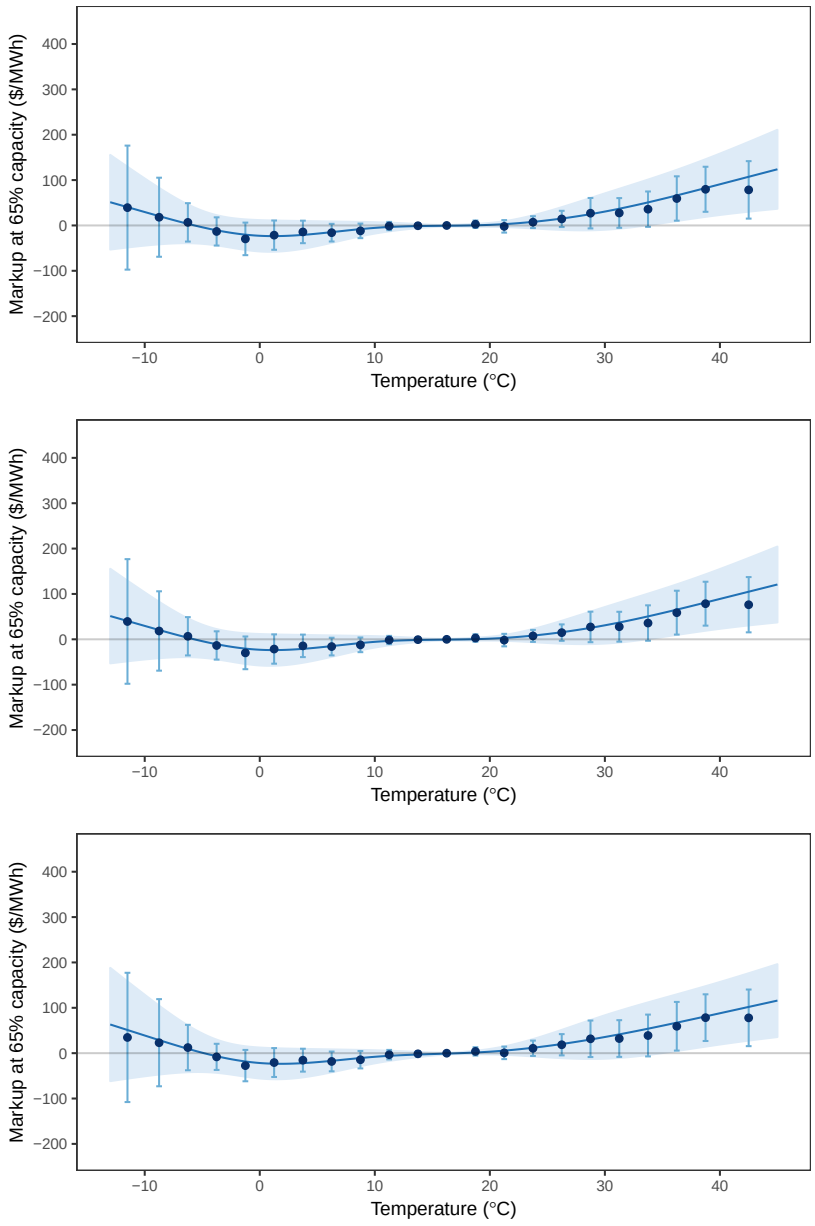
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for markup at 75% capacity. Each panel displays the results from a different regression specification; from top to bottom, the regression removes the control for local temperature, replaces the control for local temperature with a control for local capacity factor, and adds a control for local capacity factor, as compared to Equation (1.14). The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.14: Temperature-response functions for markup at 75% capacity with minimum distance



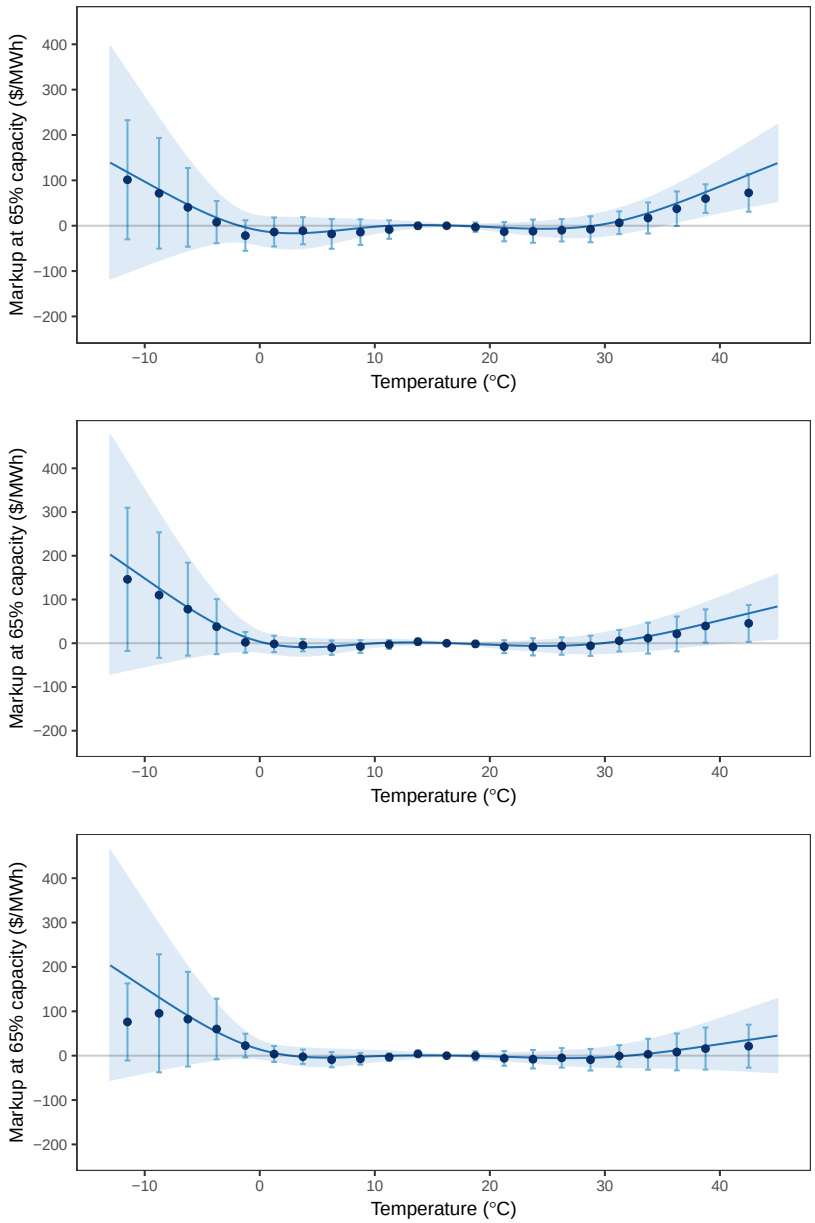
Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for markup at 75% capacity. Each panel displays the results from imposing a different minimum distance between each unit and its matched weather station; from top to bottom, the distances are 100 km, 200 km, and 300 km. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.15: Temperature-response functions for markup at 65% capacity with other local controls



Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for markup at 65% capacity. Each panel displays the results from a different regression specification; from top to bottom, the regression removes the control for local temperature, replaces the control for local temperature with a control for local capacity factor, and adds a control for local capacity factor, as compared to Equation (1.14). The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Figure A.16: Temperature-response functions for markup at 65% capacity with minimum distance



Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for markup at 65% capacity. Each panel displays the results from imposing a different minimum distance between each unit and its matched weather station; from top to bottom, the distances are 100 km, 200 km, and 300 km. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

A.4.3 Matched temperature or local temperature

Table 1.9 in the main text highlights the importance of using of a prediction model to match each power plant to the weather station that is most predictive of the plant's local market capacity, rather than simply using the local temperature closest to the power plant, when estimating the effect of local market size on markups at 85% capacity. Tables A.8–A.10 show similar results when using temperature bins in the first stage or when examining markups at 75% capacity or 65% capacity, which further highlight the importance of my prediction model and estimation procedure. Figure A.17 displays the first stages of these regressions using the local temperature, or the temperature-response functions for local market size using the local temperature. Figure A.18 displays the reduced form of these regressions using the local temperature, or the temperature-response functions for markups using the local temperature.

Table A.8: Temperature selection — Effects on markups using temperature bins

	Markup at 85% capacity ($/MWh)					
	Matched temperature			Local temperature		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−18.35**			−5.20		
	(7.63)			(4.94)		
Units in local market		−5.74**			−2.14	
		(2.79)			(2.07)	
Firms in local market			−23.82**			−6.77
			(10.64)			(7.12)
First-stage specification	Bins	Bins	Bins	Bins	Bins	Bins
First-stage F-statistic	2.93	2.11	2.63	6.28	5.67	5.74
Local temperature	X	X	X			
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variable is markup at 85% capacity. Columns (1), (2), and (3) duplicate Column (2) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results using local temperature rather than matched temperature. All regressions are estimated with 2SLS and use temperature bins in the first stage; first-stage results for Columns (4)–(6) are displayed in Figure A.17. Regressions reported in columns (1)–(3) control flexibly for local temperature; all regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.9: Temperature selection — Effects on markups at 75% capacity

	Markup at 75% capacity ($/MWh)					
	Matched temperature			Local temperature		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−15.24**			−4.70		
	(6.46)			(4.58)		
Units in local market		−5.40**			−2.12	
		(2.58)			(2.01)	
Firms in local market			−22.95**			−6.96
			(10.25)			(6.92)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	5.88	4.65	5.20	20.08	19.29	19.75
Local temperature	X	X	X			
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	5,224,152	5,224,152	5,224,152	5,224,152	5,224,152	5,224,152

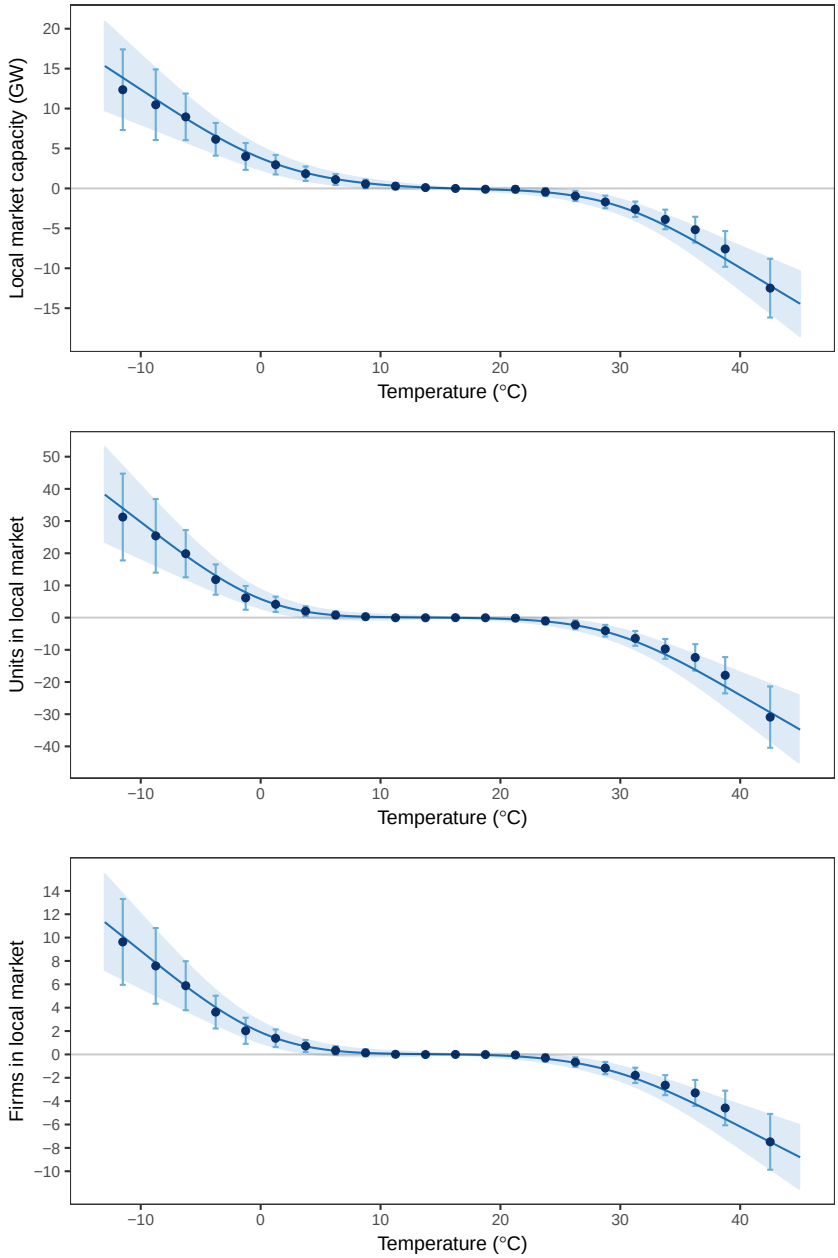
Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variable is markup at 75% capacity. Columns (1), (2), and (3) duplicate Column (6) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results using local temperature rather than matched temperature. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (4)–(6) are displayed in Figure A.17. Regressions reported in columns (1)–(3) control flexibly for local temperature; all regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.10: Temperature selection — Effects on markups at 65% capacity

	Markup at 65% capacity ($/MWh)					
	Matched temperature			Local temperature		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−13.00***			−3.44		
	(5.03)			(3.27)		
Units in local market		−5.30**			−1.47	
		(2.21)			(1.40)	
Firms in local market			−21.64**			−5.03
			(8.47)			(4.90)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	5.28	4.29	4.87	20.90	20.09	20.61
Local temperature	X	X	X			
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	5,415,331	5,415,331	5,415,331	5,415,331	5,415,331	5,415,331

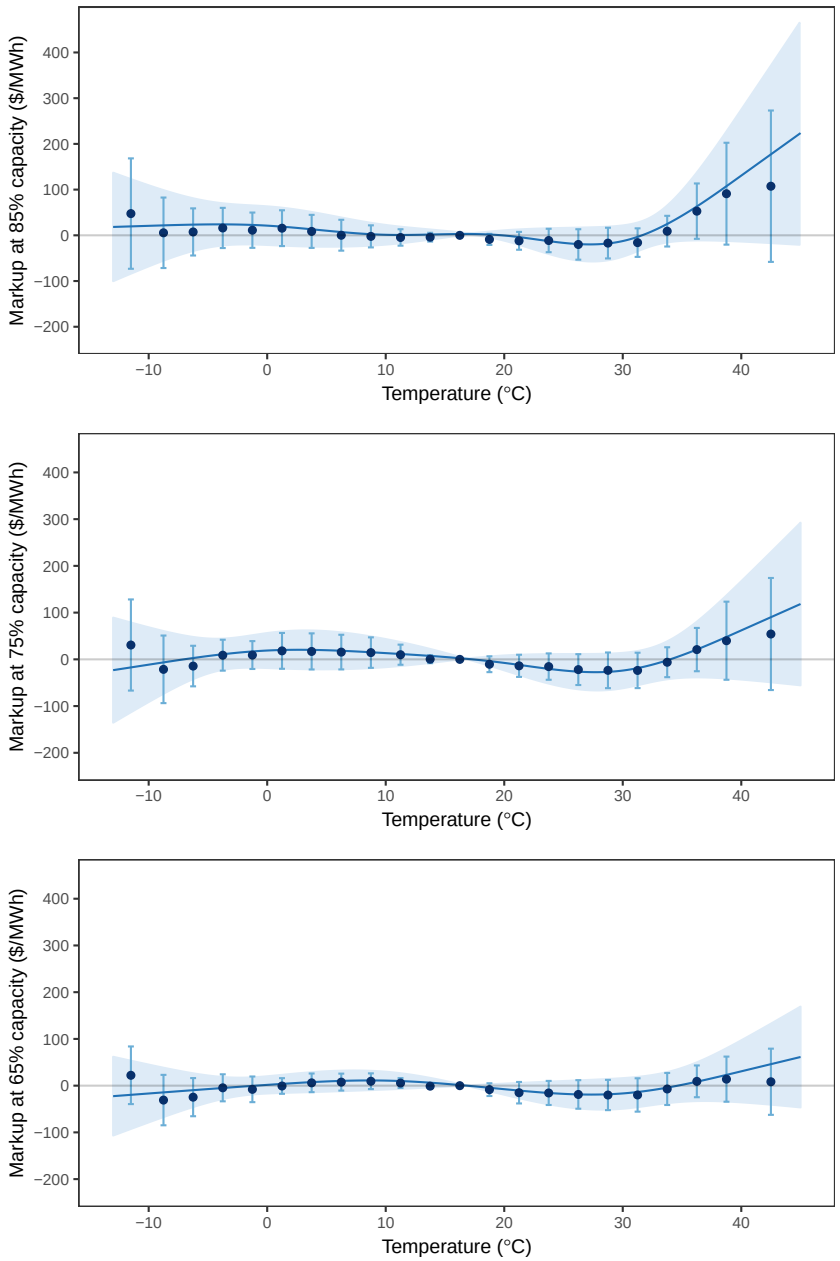
Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variable is markup at 65% capacity. Columns (1), (2), and (3) duplicate Column (9) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results using local temperature rather than matched temperature. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage; first-stage results for Columns (4)–(6) are displayed in Figure A.17. Regressions reported in columns (1)–(3) control flexibly for local temperature; all regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A.17: Temperature-response functions for market size measures using local temperature



Notes: Each panel of this figure plots a temperature-response function that results from estimating regressions similar to Equation (1.11) for a different measure of local market size. From top to bottom, the measures are local market capacity, number of units in the local market, and number of firms in the local market, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's local weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects.

Figure A.18: Temperature-response functions for markups using local temperature



Notes: Each panel of this figure plots a temperature-response function that results from estimating regressions similar to Equation (1.14) for markup at a different quantity of production. From top to bottom, the quantities are 85%, 75%, and 65%, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's local weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions include unit-by-hour-by-month fixed effects and hour of sample fixed effects.

A.4.4 Offer prices

Table A.11 displays the results of regressions that estimate the effect of local market size on offer prices at 85% capacity, rather than markups at 85% capacity. The results for both sets of regressions are nearly identical, suggesting that the effect of local market size on markups is entirely due to firms strategically changing offer prices and not due to changes in marginal cost. Tables A.12–A.14 show similar results when using temperature bins in the first stage or when examining offer prices at 75% capacity or 65% capacity, which further confirm that the effects on markups are due to strategic firm behavior. Figure A.19 displays the reduced form of these regressions with offer price as the dependent variable, or the temperature-response functions for offer prices.

Table A.11: Robustness to outcome measure — Effects on offer prices at 85% capacity

	Markup at 85% capacity ($/MWh)			Offer price at 85% capacity ($/MWh)		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−25.42**			−25.19**		
	(11.22)			(11.19)		
Units in local market		−10.27**			−10.21**	
		(5.06)			(5.05)	
Firms in local market			−41.10**			−40.89**
			(19.17)			(19.14)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	6.47	4.98	5.44	6.47	4.98	5.44
Local temperature	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variables are markup at 85% capacity and offer price at 85% capacity, as shown in the column headers. Columns (1), (2), and (3) duplicate Column (3) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results with offer price as the dependent variable. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.12: Robustness to outcome measure — Effects on offer prices using temperature bins

	Markup at 85% capacity ($/MWh)			Offer price at 85% capacity ($/MWh)		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−18.35**			−18.21**		
	(7.63)			(7.62)		
Units in local market		−5.74**			−5.71**	
		(2.79)			(2.79)	
Firms in local market			−23.82**			−23.73**
			(10.64)			(10.63)
First-stage specification	Bins	Bins	Bins	Bins	Bins	Bins
First-stage F-statistic	2.93	2.11	2.63	2.93	2.11	2.63
Local temperature	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775	4,820,775

Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variables are markup at 85% capacity and offer price at 85% capacity, as shown in the column headers. Columns (1), (2), and (3) duplicate Column (2) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results with offer price as the dependent variable. All regressions are estimated with 2SLS and use temperature bins in the first stage. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.13: Robustness to outcome measure — Effects on offer prices at 75% capacity

	Markup at 75% capacity ($/MWh)			Offer price at 75% capacity ($/MWh)		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−15.24**			−14.99**		
	(6.46)			(6.45)		
Units in local market		−5.40**			−5.35**	
		(2.58)			(2.57)	
Firms in local market			−22.95**			−22.74**
			(10.25)			(10.24)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	5.88	4.65	5.20	5.88	4.65	5.20
Local temperature	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	5,224,152	5,224,152	5,224,152	5,224,152	5,224,152	5,224,152

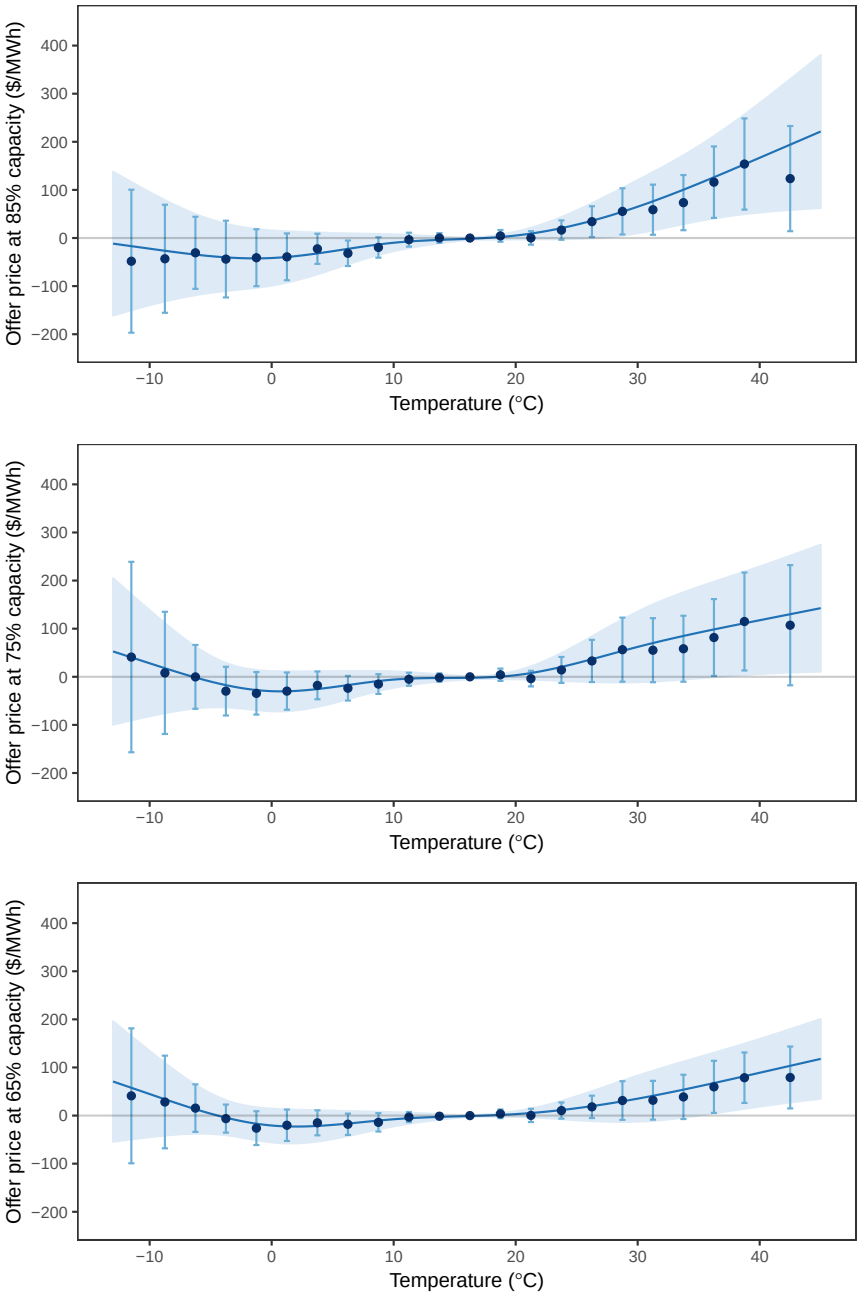
Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variables are markup at 75% capacity and offer price at 75% capacity, as shown in the column headers. Columns (1), (2), and (3) duplicate Column (6) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results with offer price as the dependent variable. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.14: Robustness to outcome measure — Effects on offer prices at 65% capacity

	Markup at 65% capacity ($/MWh)			Offer price at 65% capacity ($/MWh)		
	(1)	(2)	(3)	(4)	(5)	(6)
Local market capacity (GW)	−13.00***			−12.72**		
	(5.03)			(5.01)		
Units in local market		−5.30**			−5.25**	
		(2.21)			(2.20)	
Firms in local market			−21.64**			−21.45**
			(8.47)			(8.47)
First-stage specification	Spline	Spline	Spline	Spline	Spline	Spline
First-stage F-statistic	5.28	4.29	4.87	5.28	4.29	4.87
Local temperature	X	X	X	X	X	X
Unit × hour × month FEs	X	X	X	X	X	X
Hour of sample FEs	X	X	X	X	X	X
Observations	5,415,331	5,415,331	5,415,331	5,415,331	5,415,331	5,415,331

Notes: This table reports results of estimating regressions similar to Equation (1.10) with different measures of local market size as the independent variable. The dependent variables are markup at 65% capacity and offer price at 65% capacity, as shown in the column headers. Columns (1), (2), and (3) duplicate Column (9) of Tables 1.3, 1.4, and 1.5, respectively. The remaining columns report results with offer price as the dependent variable. All regressions are estimated with 2SLS and use a natural cubic spline in the first stage. All regressions control flexibly for local temperature and include unit-by-hour-by-month fixed effects and hour of sample fixed effects. Standard errors are in parentheses and are two-way clustered at the unit and date. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A.19: Temperature-response functions for offer prices

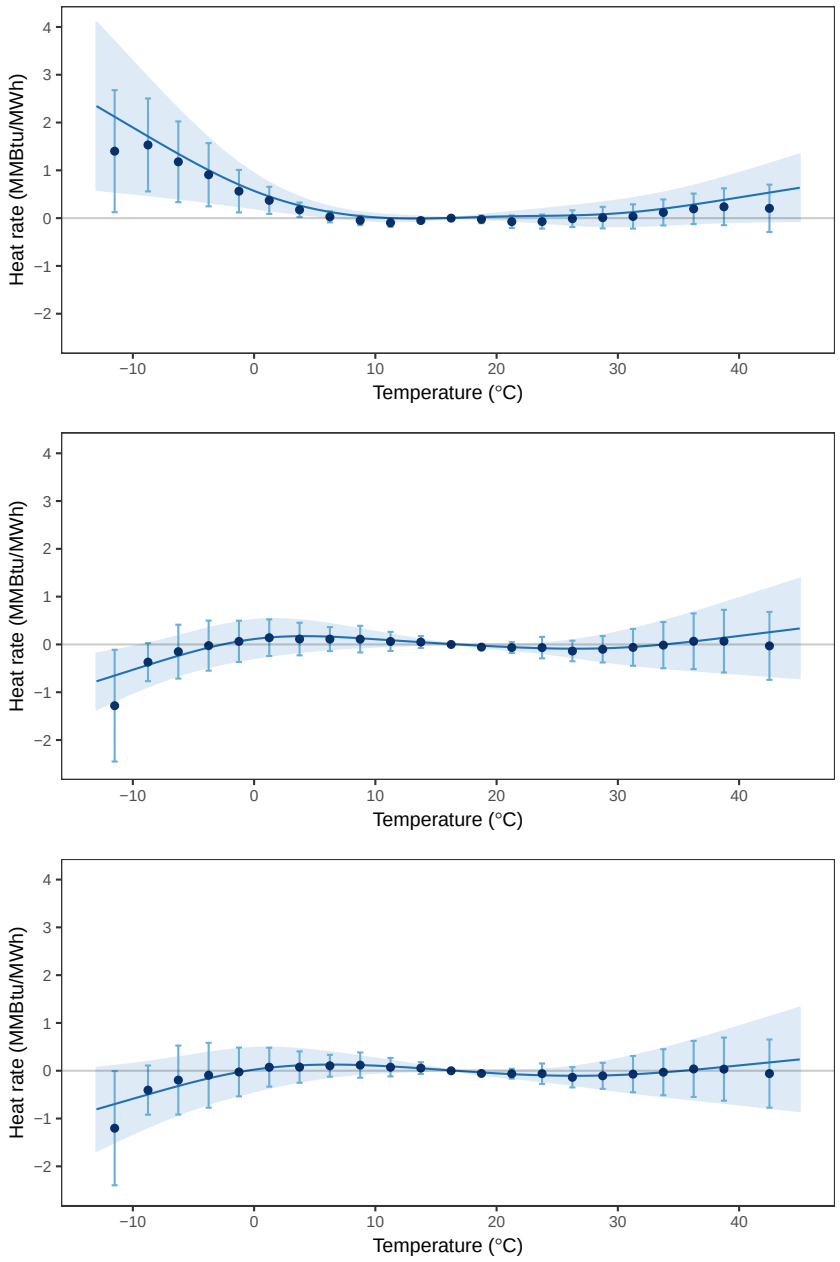


Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for offer price at a different quantity of production. From top to bottom, the quantities are 85%, 75%, and 65%, as denoted on the y-axis of each panel. The x-axis in each panel is the temperature at each unit's matched weather station. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

A.4.5 Heat rate

Figure A.20 displays temperature-response functions for a unit's heat rate, which is an inverse measure of the unit's efficiency. The top panel plots the temperature-response function using the unit's local temperature, the middle panel plots the temperature-response function using the unit's matched temperature, and the bottom panel plots the temperature-response function using the unit's matched temperature while controlling for the unit's local temperature. This figure shows that neither extreme heat at a unit's local weather station nor extreme heat at a unit's matched weather station substantially affects the unit's efficiency. Thus, the main results of this analysis cannot be explained by temperature effects on unit efficiency.

Figure A.20: Temperature-response function for heat rate



Notes: Each panel of this figure plots a temperature-response function that results from estimating a regression similar to Equation (1.14) for heat rate. Each panel displays the results from a different regression specification; from top to bottom, the regression uses a unit's local temperature, uses a unit's matched temperature, and uses a unit's matched temperature while controlling for the unit's local temperature. The x-axis in each panel is temperature, either at each unit's local weather station or at each unit's matched weather station, as described above. The points and error bars show estimates and 95% confidence intervals for a nonparametric function of temperature using 2.5°C (4.5°F) temperature bins; the points are at the midpoints of the bins. The curve and shaded area show the estimate and 95% confidence interval for a natural cubic spline of temperature with knots at 10°C (18°F) increments between -10°C (14°F) and 40°C (104°F). Estimates are relative to a temperature of 16.25°C (61.25°F). Confidence intervals use standard errors that are two-way clustered at the unit and date. Regressions control for a nonparametric function of local temperature, unit-by-hour-by-month fixed effects, and hour of sample fixed effects.

Appendix B

Appendix: Panel Data and Experimental Design

B.1 Derivations and proofs

We are the first to derive analytic power calculation formulas for panel data models under non-i.i.d. error structures, accounting for (i) heterogeneous error structures across panel units, and (ii) random time shocks. In this section, we derive power calculation analytics for cross-sectional, difference-in-differences (DD), collapsed data, and analysis of covariance (ANCOVA) estimators. We present the resulting equations in Section 3.2 of the main text.

We first re-derive well-known power calculation formulas, for both a cross-sectional experiment and a panel experiment that applies the difference-in-differences estimator, under the assumption that error terms are uncorrelated. We then relax this assumption to consider the difference-in-differences estimator applied to a panel experiment in which the error structure of the data exhibits an arbitrary form of serial correlation. This resembles the most general model derived by Frison and Pocock (1992), while incorporating two features characteristic of panel data in economics: (i) error structures that vary across individual panel units; and (ii) random time series shocks that are common to all panel units. We then consider a collapsed data research design, and demonstrate that collapsing to a single pre/post observation for each panel unit *ex post* does not obviate the need to account for non-constant within-unit serial correlation *ex ante*. We also conduct a parallel series of derivations using the ANCOVA estimator, which underscores the strong assumptions required to achieve analytical tractability in this regression model.

Next, we prove Lemma 1 presented in the main text. We also prove an analogous Lemma B1 for a cross-sectional model. These lemmas show that the above power calculation formulas can be applied even in the presence of cross-sectional correlations, so long as treatment is randomly assigned at the unit level. In particular, Lemma 1 states that the variance estimator we derive, which accounts for non-constant serial correlation, gives an unbiased estimate of the *ex ante* expected variance under unit-level randomization, even when cross-sectional correlations exist. In other words, our newly derived power calculation formula can

be applied to any panel experiment setting, regardless of the true error structure of the data, so long as treatment is randomly assigned to units.

Finally, we demonstrate that replacing unit and/or time fixed effects with a Treat_i and/or Post_t dummy, respectively, does not alter the statistical power of the DD estimator, as long as researchers implement the CRVE *ex post*. This analysis complements our simulation results presented in Section 3.2.2.2 of the main text.

B.1.1 Cross sectional experiments

Model There are J units randomly assigned a treatment status D_i, with proportion P in treatment ($D_i = 1$) and proportion $(1 - P)$ in control ($D_i = 0$). The units are indexed so $i \in [1, PJ]$ is treated and $j \in [PJ + 1, J]$ is a control. We make standard assumptions for randomized trials:

Assumption B1 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_i = \beta + \tau D_i + \varepsilon_i \tag{B.1}$$

where Y_i is the outcome of interest for unit i; β is the expected outcome of non-treated units; τ is the treatment effect which is homogeneous across all units; D_i is a treatment indicator; and ε_i is an idiosyncratic error term distributed i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Assumption B2 (Strict exogeneity)**.** $\mathbb{E}[\varepsilon_i \mid \mathbf{X}] = 0$*, where* $\mathbf{X} = [\mathbf{1}\ \mathbf{D}]$*. In practice, this follows from random assignment of D_i.*

Coefficient estimate The coefficient estimates from an OLS regression are

$$\begin{aligned}
\begin{pmatrix} \widehat{\beta} \\ \widehat{\tau} \end{pmatrix} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \begin{pmatrix} J & PJ \\ PJ & PJ \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{J} Y_i \\ \sum_{i=1}^{J} D_i Y_i \end{pmatrix} \\
&= \frac{1}{P(1-P)J^2} \begin{pmatrix} PJ\left(\sum_{i=1}^{J} Y_i - \sum_{i=1}^{J} D_i Y_i\right) \\ J\left(\sum_{i=1}^{J} D_i Y_i - P\sum_{i=1}^{J} Y_i\right) \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{(1-P)J}\sum_{i=PJ+1}^{J} Y_i \\ \frac{1}{PJ}\sum_{i=1}^{PJ} Y_i - \frac{1}{(1-P)J}\sum_{i=PJ+1}^{J} Y_i \end{pmatrix}
\end{aligned}$$

Defining the mean outcome in the treatment and control groups, respectively, as

$$\overline{Y}_T = \frac{1}{PJ}\sum_{i=1}^{PJ} Y_i$$

$$\overline{Y}_C = \frac{1}{(1-P)J}\sum_{i=PJ+1}^{J} Y_i$$

gives coefficient estimates of

$$\begin{aligned}\widehat{\beta} &= \overline{Y}_C \\ \widehat{\tau} &= \overline{Y}_T - \overline{Y}_C\end{aligned}$$

Variance of coefficient estimate The variance of the estimate of the treatment effect, $\widehat{\tau}$, is

$$\begin{aligned}\operatorname{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= \operatorname{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{Y}_T, \overline{Y}_C \mid \mathbf{X}\right) \\ &= \operatorname{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) \qquad \text{(B.2)}\end{aligned}$$

The first term of Equation (B.2) is

$$\begin{aligned}\operatorname{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) &= \operatorname{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} Y_i \mid \mathbf{X}\right) \\ &= \frac{\sigma_\varepsilon^2}{PJ} \qquad \text{(B.3)}\end{aligned}$$

Similarly, the second term of Equation (B.2) is

$$\operatorname{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) = \frac{\sigma_\varepsilon^2}{(1-P)J} \qquad \text{(B.4)}$$

Substituting Equations (B.3) and (B.4) into Equation (B.2) gives

$$\begin{aligned}\operatorname{Var}(\widehat{\tau} \mid \mathbf{X}) &= \frac{\sigma_\varepsilon^2}{PJ} + \frac{\sigma_\varepsilon^2}{(1-P)J} \\ &= \frac{\sigma_\varepsilon^2}{P(1-P)J} \qquad \text{(B.5)}\end{aligned}$$

The variance estimator produced by an OLS regression of Equation (B.1) is an unbiased estimator of this variance.

Minimum detectable effect The minimum detectable effect (MDE), or the smallest treatment effect we have the power to detect, is

$$\begin{aligned}MDE &= \left(t_{1-\kappa}^{J-2} + t_{\alpha/2}^{J-2}\right)\sqrt{\operatorname{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\ &= \left(t_{1-\kappa}^{J-2} + t_{\alpha/2}^{J-2}\right)\sqrt{\frac{\sigma_\varepsilon^2}{P(1-P)J}} \qquad \text{(B.6)}\end{aligned}$$

where κ is the power of the hypothesis test, α is the significance level, and the critical values are drawn from t-distributions with $J-2$ degrees of freedom.

B.1.2 Panel experiments

B.1.2.1 Independent error structure

Model In this model, P proportion of the J units are again randomized into treatment. The researcher collects the outcome Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For units in the treatment group, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for units in the control group, $D_{it} = 0$ in all $(m + r)$ periods.

Assumption B3 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it} \tag{B.7}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed i.i.d. $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption B4 (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J - 1$ *unit dummies, and* $(m + r) - 1$ *time dummies. This again follows from random assignment of* D_{it}*.*

Assumption B5 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Coefficient estimate The treatment effect, τ, can be estimated by OLS with unit and time fixed effects. In a balanced panel, this is equivalent to de-meaning at both the unit and time levels. Define

$$\ddot{Y}_{it} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{\overline{Y}} \tag{B.8}$$

$$\ddot{D}_{it} = D_{it} - \overline{D}_i - \overline{D}_t + \overline{\overline{D}} \tag{B.9}$$

$$\ddot{\omega}_{it} = \omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}} \tag{B.10}$$

where

$$\overline{Y}_i = \frac{1}{m + r} \sum_{t=-m+1}^{r} Y_{it}$$

$$\overline{Y}_t = \frac{1}{J} \sum_{i=1}^{J} Y_{it}$$

$$\overline{\overline{Y}} = \frac{1}{J(m + r)} \sum_{t=-m+1}^{r} \sum_{i=1}^{J} Y_{it}$$

with $\overline{D}_i$, $\overline{D}_t$, $\overline{\overline{D}}$, $\overline{\omega}_i$, $\overline{\omega}_t$, and $\overline{\overline{\omega}}$ defined analogously. Substituting Equations (B.8)–(B.10) into Equation (B.7) and simplifying gives the de-meaned DGP

$$\ddot{Y}_{it} = \tau \ddot{D}_{it} + \ddot{\omega}_{it}$$

The estimate of the treatment effect is

$$\begin{aligned}
\widehat{\tau} &= (\ddot{\mathbf{D}}'\ddot{\mathbf{D}})^{-1}\ddot{\mathbf{D}}'\ddot{\mathbf{Y}} \\
&= \left(\sum_{i=1}^{J}\sum_{t=-m+1}^{r} \ddot{D}_{it}^2\right)^{-1} \sum_{i=1}^{J}\sum_{t=-m+1}^{r} \ddot{D}_{it}\ddot{Y}_{it} \\
&= \frac{m+r}{P(1-P)Jmr}\left[\sum_{i=1}^{J}\sum_{t=-m+1}^{r} \ddot{D}_{it}Y_{it} - \sum_{i=1}^{J}\overline{Y}_i \sum_{t=-m+1}^{r} \ddot{D}_{it}\right. \\
&\qquad \left. - \sum_{t=-m+1}^{r}\overline{Y}_t \sum_{i=1}^{J} \ddot{D}_{it} + \overline{Y}\sum_{i=1}^{J}\sum_{t=-m+1}^{r} \ddot{D}_{it}\right] \\
&= \frac{m+r}{P(1-P)Jmr}\sum_{i=1}^{J}\sum_{t=-m+1}^{r} \ddot{D}_{it}Y_{it} \\
&= \frac{m+r}{P(1-P)Jmr}\left[\sum_{i=1}^{PJ}\left(\frac{(1-P)m}{m+r}\sum_{t=1}^{r} Y_{it} - \frac{(1-P)r}{m+r}\sum_{t=-m+1}^{0} Y_{it}\right)\right. \\
&\qquad \left. + \sum_{i=PJ+1}^{J}\left(\frac{Pr}{m+r}\sum_{t=-m+1}^{0} Y_{it} - \frac{Pm}{m+r}\sum_{t=1}^{r} Y_{it}\right)\right] \\
&= \left(\overline{Y}_T^A - \overline{Y}_T^B\right) - \left(\overline{Y}_C^A - \overline{Y}_C^B\right) \qquad \text{(B.11)}
\end{aligned}$$

where

$$\begin{aligned}
\overline{Y}_T^A &= \frac{1}{PJr}\sum_{i=1}^{PJ}\sum_{t=1}^{r} Y_{it} \\
\overline{Y}_T^B &= \frac{1}{PJm}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0} Y_{it} \\
\overline{Y}_C^A &= \frac{1}{(1-P)Jr}\sum_{i=PJ+1}^{J}\sum_{t=1}^{r} Y_{it} \\
\overline{Y}_C^B &= \frac{1}{(1-P)Jm}\sum_{i=PJ+1}^{J}\sum_{t=-m+1}^{0} Y_{it}
\end{aligned}$$

Variance of coefficient estimate The variance of the estimate of the treatment effect, $\widehat{\tau}$, is

$$\text{Var}(\widehat{\tau} \mid \mathbf{X}) = \text{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) + \text{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) + \text{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) + \text{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right)$$

$$
\begin{aligned}
&- 2\,\mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) - 2\,\mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) + 2\,\mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^B \mid \mathbf{X}\right) \\
&+ 2\,\mathrm{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^A \mid \mathbf{X}\right) - 2\,\mathrm{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) - 2\,\mathrm{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) \\
&= \mathrm{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) + \mathrm{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) + \mathrm{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) + \mathrm{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) \\
&- 2\Bigg[\,\mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) + \mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) \\
&\quad + \mathrm{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) + \mathrm{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right)\Bigg]
\end{aligned} \tag{B.12}
$$

The first term of Equation (B.12) is

$$
\begin{aligned}
\mathrm{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) &= \mathrm{Var}\left(\frac{1}{PJr}\sum_{i=1}^{PJ}\sum_{t=1}^{r} Y_{it} \mid \mathbf{X}\right) \\
&= \mathrm{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} \upsilon_i \mid \mathbf{X}\right) + \mathrm{Var}\left(\frac{1}{r}\sum_{t=1}^{r} \delta_t \mid \mathbf{X}\right) \\
&\quad + \mathrm{Var}\left(\frac{1}{PJr}\sum_{i=1}^{PJ}\sum_{t=1}^{r} \omega_{it} \mid \mathbf{X}\right) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right)
\end{aligned} \tag{B.13}
$$

Similarly, the remaining variance terms of Equation (B.12) are

$$
\mathrm{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) = \frac{1}{PJm}\left(m\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right) \tag{B.14}
$$

$$
\mathrm{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) = \frac{1}{(1-P)Jr}\left(r\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2\right) \tag{B.15}
$$

$$
\mathrm{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)Jm}\left(m\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2\right) \tag{B.16}
$$

The first covariance component of Equation (B.12) is

$$
\begin{aligned}
\mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) &= \mathbb{E}\left[\overline{Y}_T^A \overline{Y}_T^B \mid \mathbf{X}\right] - \mathbb{E}\left[\overline{Y}_T^A \mid \mathbf{X}\right]\mathbb{E}\left[\overline{Y}_T^B \mid \mathbf{X}\right] \\
&= \beta(\beta+\tau) + \frac{1}{PJ}\,\mathbb{E}\left[\upsilon_i^2 \mid \mathbf{X}\right] - \beta(\beta+\tau) \\
&= \frac{\sigma_\upsilon^2}{PJ}
\end{aligned} \tag{B.17}
$$

Similarly, the remaining covariance terms of Equation (B.12) are

$$
\mathrm{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{r} \tag{B.18}
$$

$$\text{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{m} \tag{B.19}$$

$$\text{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{\sigma_\upsilon^2}{(1-P)J} \tag{B.20}$$

Substituting Equations (B.13)–(B.20) into Equation (B.12) gives

$$\begin{aligned}\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= 2\frac{\sigma_\upsilon^2}{P(1-P)J} + 2\frac{(m+r)\sigma_\delta^2}{mr} + \frac{(m+r)\sigma_\omega^2}{P(1-P)Jmr} - 2\left[\frac{\sigma_\upsilon^2}{P(1-P)J} + \frac{(m+r)\sigma_\delta^2}{mr}\right] \\ &= \left(\frac{\sigma_\omega^2}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right) \qquad \text{(B.21)}\end{aligned}$$

The variance estimator produced by an OLS regression of Equation (B.7) is an unbiased estimator of this variance.

Minimum detectable effect The MDE is

$$\begin{aligned}MDE &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\ &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\left(\frac{\sigma_\omega^2}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)}\end{aligned}$$

This is the standard Frison and Pocock (1992) result (page 1693), also referenced by McKenzie (2012), and is shown as Equation (3.3) in the main text. We assume J degrees of freedom to be consistent with the assumptions of the CRVE; alternatively, $J(m+r) - (J+m+r)$ degrees of freedom would be consistent with the assumptions of OLS standard errors. Note that this model differs slightly from Frison and Pocock (1992) and McKenzie (2012), which both assume $\sigma_\delta^2 = 0$.

B.1.2.2 Serially correlated error structure

Model There are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption B6 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it} \tag{B.22}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption B7 (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J-1$ *unit dummies, and* $(m+r)-1$ *time dummies. This again follows from random assignment of* D_{it}*.*

Assumption B8 (Balanced panel)**.** *The number of pre-treatment observations,* m*, and post-treatment observations,* r*, is the same for each unit, and all units are observed in every time period.*

Assumption B9 (Independence across units)**.** $\mathbb{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ \forall\ t, s.$

Assumption B10 (Symmetric covariance structures)**.** *Define:*

$$\psi^B \equiv \frac{2}{Jm(m-1)} \sum_{i=1}^{J} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^A \equiv \frac{2}{Jr(r-1)} \sum_{i=1}^{J} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^X \equiv \frac{1}{Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of the same unit, respectively. Define ψ_T^B*,* ψ_T^A*, and* ψ_C^X *analogously, where we consider only the* PJ *treated units; also define* ψ_C^B*,* ψ_C^A*, and* ψ_C^X *analogously, where we consider only the* $(1-P)J$ *control units. Using these definitions, assume that* $\psi^B = \psi_T^B = \psi_C^B$*;* $\psi^A = \psi_T^A = \psi_C^A$*; and* $\psi^X = \psi_T^X = \psi_C^X$*.*[1]

Coefficient estimate The treatment effect estimator is the same as Equation (B.11):

$$\widehat{\tau} = \left(\overline{Y}_T^A - \overline{Y}_T^B\right) - \left(\overline{Y}_C^A - \overline{Y}_C^B\right)$$

Variance of coefficient estimate With a serially correlated error structure, Equation (B.12) is still correct by Assumption B9, but Equations (B.13)–(B.20) no longer hold. With serially correlated errors, the first term of Equation (B.12) is

$$\begin{aligned}\text{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) &= \text{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it} \mid \mathbf{X}\right) \\ &= \text{Var}\left(\frac{1}{PJ} \sum_{i=1}^{PJ} \upsilon_i \mid \mathbf{X}\right) + \text{Var}\left(\frac{1}{r} \sum_{t=1}^{r} \delta_t \mid \mathbf{X}\right)\end{aligned}$$

[1] We choose the letters "B" to indicate the Before-treatment period, and "A" to indicate the After-treatment period. We index the m pre-treatment periods $\{-m+1, \ldots, 0\}$, and the r post-treatment periods $\{1, \ldots, r\}$. In a randomized setting, $\mathbb{E}\left[\psi^B\right] = \mathbb{E}\left[\psi_T^B\right] = \mathbb{E}\left[\psi_C^B\right]$, $\mathbb{E}\left[\psi^A\right] = \mathbb{E}\left[\psi_T^A\right] = \mathbb{E}\left[\psi_C^A\right]$, and $\mathbb{E}\left[\psi^X\right] = \mathbb{E}\left[\psi_T^X\right] = \mathbb{E}\left[\psi_C^X\right]$, making this a reasonable assumption *ex ante*. However, it is possible for treatment to alter the covariance structure of treated units only.

$$
\begin{aligned}
&\quad + \operatorname{Var}\left(\frac{1}{PJr}\sum_{i=1}^{PJ}\sum_{t=1}^{r}\omega_{it} \mid \mathbf{X}\right) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right) \\
&\quad + \frac{2}{(PJr)^2}\sum_{i=1}^{PJ}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it},\omega_{is} \mid \mathbf{X}) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi^A\right) \qquad \text{(B.23)}
\end{aligned}
$$

Similarly, with serial correlation, the remaining variance terms of Equation (B.12) are

$$
\operatorname{Var}\left(\overline{Y}_T^B \mid \mathbf{X}\right) = \frac{1}{PJm}\left(m\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi^B\right) \qquad \text{(B.24)}
$$

$$
\operatorname{Var}\left(\overline{Y}_C^A \mid \mathbf{X}\right) = \frac{1}{(1-P)Jr}\left(r\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi^A\right) \qquad \text{(B.25)}
$$

$$
\operatorname{Var}\left(\overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)Jm}\left(m\sigma_\upsilon^2 + (1-P)J\sigma_\delta^2 + \sigma_\omega^2 + (m-1)\psi^B\right) \qquad \text{(B.26)}
$$

With serial correlation, the first covariance component of Equation (B.12) is

$$
\begin{aligned}
\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_T^B \mid \mathbf{X}\right) &= \mathbb{E}\left[\overline{Y}_T^A\overline{Y}_T^B \mid \mathbf{X}\right] - \mathbb{E}\left[\overline{Y}_T^A \mid \mathbf{X}\right]\mathbb{E}\left[\overline{Y}_T^B \mid \mathbf{X}\right] \\
&= \beta(\beta+\tau) + \frac{1}{PJ}\mathbb{E}\left[\upsilon_i^2 \mid \mathbf{X}\right] \\
&\quad + \frac{1}{(PJ)^2mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathbb{E}[\omega_{it}\omega_{is} \mid \mathbf{X}] - \beta(\beta+\tau) \\
&= \frac{\sigma_\upsilon^2}{PJ} + \frac{1}{(PJ)^2mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\operatorname{Cov}(\omega_{it},\omega_{is} \mid \mathbf{X}) \\
&= \frac{1}{PJ}\left(\sigma_\upsilon^2 + \psi^X\right) \qquad \text{(B.27)}
\end{aligned}
$$

Similarly, with serial correlation, the remaining covariance terms of Equation (B.12) are

$$
\operatorname{Cov}\left(\overline{Y}_T^A, \overline{Y}_C^A \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{r} \qquad \text{(B.28)}
$$

$$
\operatorname{Cov}\left(\overline{Y}_T^B, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{\sigma_\delta^2}{m} \qquad \text{(B.29)}
$$

$$
\operatorname{Cov}\left(\overline{Y}_C^A, \overline{Y}_C^B \mid \mathbf{X}\right) = \frac{1}{(1-P)J}\left(\sigma_\upsilon^2 + \psi^X\right) \qquad \text{(B.30)}
$$

Substituting Equations (B.23)–(B.30) into Equation (B.12) and simplifying gives

$$
\operatorname{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) = \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2 + (r-1)\psi^A\right)
$$

$$
\begin{aligned}
&+ \frac{1}{PJm}\left(m\sigma_{\upsilon}^2 + PJ\sigma_{\delta}^2 + \sigma_{\omega}^2 + (m-1)\psi^B\right) \\
&+ \frac{1}{(1-P)Jr}\left(r\sigma_{\upsilon}^2 + (1-P)J\sigma_{\delta}^2 + \sigma_{\omega}^2 + (r-1)\psi^A\right) \\
&+ \frac{1}{(1-P)Jm}\left(m\sigma_{\upsilon}^2 + (1-P)J\sigma_{\delta}^2 + \sigma_{\omega}^2 + (m-1)\psi^B\right) \\
&- 2\left[\frac{1}{PJ}\left(\sigma_{\upsilon}^2 + \psi^X\right) + \frac{\sigma_{\delta}^2}{r} + \frac{\sigma_{\delta}^2}{m} + \frac{1}{(1-P)J}\left(\sigma_{\upsilon}^2 + \psi^X\right)\right] \\
&= \left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right] \qquad \text{(B.31)}
\end{aligned}
$$

Equation (B.31) differs from Equation (B.21), the variance of $\widehat{\tau}$ that is estimated by an OLS regression, due to the presence of the ψ terms that describe the serially correlated error structure. This error structure alters the true variance of the treatment effect estimator such that the OLS estimator of the variance is not correct. The cluster-robust variance estimator must be used for correct inference.

Minimum detectable effect With serial correlation, the MDE is

$$
\begin{aligned}
MDE &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\
&= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}
\end{aligned}
$$

This is the serial-correlation-robust (SCR) power calculation formula, found in Equation (3.2) in the main text. Note that ψ^B, ψ^A, and ψ^X are likely to depend on the length of the pre- and post- treatment periods, m and r; serial correlation often diminishes as time periods become further apart, so larger values of m and r will result in less correlation on average and smaller parameter values. These parameters do not depend in a systematic way on the number of experimental units, J, however.

B.1.2.3 Serially and cross-sectionally correlated error structure

Randomization at the unit level justifies Assumptions B9 and B10, but we can also characterize the full variance without this random assignment assumption. This highlights the type of correlation structures that a researcher might face when wanting to perform a power calculation for a quasi-experimental design.

Model There are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption B11 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption B12 (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$, *where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$, $J-1$ *unit dummies, and* $(m+r)-1$ *time dummies. This again follows from random assignment of* D_{it}.

Assumption B13 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption B14 (Independence across units at different times)**.** $\mathbb{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0, \ \forall \ i \neq j, \ t \neq s$.

Define ψ_i to be the average serial correlation parameters previously defined in Appendix B.1.2.2, with the subscript i denoting the correlation is within-unit serial correlation. Also define ψ_t to be the comparable parameters characterizing cross-sectional correlations, with the subscript t denoting the correlation is cross-sectional within a time period. For example, the average cross-sectional covariance among the treated group post-treatment is

$$\psi_{t,T}^A = \frac{2}{PJ(PJ-1)r} \sum_{i=1}^{PJ-1} \sum_{j=i+1}^{PJ} \sum_{t=1}^{r} \text{Cov}(\omega_{it}, \omega_{jt} \mid \mathbf{X})$$

Coefficient estimate The treatment effect estimator is the same as Equation (B.11):

$$\widehat{\tau} = \left(\overline{Y}_T^A - \overline{Y}_T^B\right) - \left(\overline{Y}_C^A - \overline{Y}_C^B\right)$$

Variance of coefficient estimate As with the serially correlated error structure in Appendix B.1.2.2, Equation (B.12) is still correct, in this case by Assumption B14. However, neither Equations (B.13)–(B.20) nor Equations (B.23)–(B.30) hold. With these arbitrary correlations, the first term of Equation (B.12) is

$$\begin{aligned}
\text{Var}\left(\overline{Y}_T^A \mid \mathbf{X}\right) &= \text{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it} \mid \mathbf{X}\right) \\
&= \text{Var}\left(\frac{1}{PJ} \sum_{i=1}^{PJ} \upsilon_i \mid \mathbf{X}\right) + \text{Var}\left(\frac{1}{r} \sum_{t=1}^{r} \delta_t \mid \mathbf{X}\right) \\
&\quad + \text{Var}\left(\frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} \omega_{it} \mid \mathbf{X}\right) \\
&= \frac{1}{PJr}\left(r\sigma_\upsilon^2 + PJ\sigma_\delta^2 + \sigma_\omega^2\right)
\end{aligned}$$

$$+\frac{2}{(PJr)^2}\sum_{i=1}^{PJ}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\mathrm{Cov}(\omega_{it},\omega_{is}\mid\mathbf{X})$$
$$+\frac{2}{(PJr)^2}\sum_{t=1}^{r}\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ}\mathrm{Cov}(\omega_{it},\omega_{jt}\mid\mathbf{X})$$
$$=\frac{1}{PJr}\left(r\sigma_v^2+PJ\sigma_\delta^2+\sigma_\omega^2+(r-1)\psi_{i,T}^A+(PJ-1)\psi_{t,T}^A\right) \tag{B.32}$$

Similarly, with arbitrary correlations, the remaining variance terms of Equation (B.12) are

$$\mathrm{Var}\left(\overline{Y}_T^B\mid\mathbf{X}\right)=\frac{1}{PJm}\left(m\sigma_v^2+PJ\sigma_\delta^2+\sigma_\omega^2+(m-1)\psi_{i,T}^B+(PJ-1)\psi_{t,T}^B\right) \tag{B.33}$$

$$\mathrm{Var}\left(\overline{Y}_C^A\mid\mathbf{X}\right)=\frac{1}{(1-P)Jr}\left(r\sigma_v^2+(1-P)J\sigma_\delta^2+\sigma_\omega^2\right.$$
$$\left.+(r-1)\psi_{i,C}^A+((1-P)J-1)\psi_{t,C}^A\right) \tag{B.34}$$

$$\mathrm{Var}\left(\overline{Y}_C^A\mid\mathbf{X}\right)=\frac{1}{(1-P)Jm}\left(m\sigma_v^2+(1-P)J\sigma_\delta^2+\sigma_\omega^2\right.$$
$$\left.+(m-1)\psi_{i,C}^B+((1-P)J-1)\psi_{t,C}^B\right) \tag{B.35}$$

With arbitrary correlations, the first covariance component of Equation (B.12) is

$$\begin{aligned}\mathrm{Cov}\left(\overline{Y}_T^A,\overline{Y}_T^B\mid\mathbf{X}\right)&=\mathbb{E}\left[\overline{Y}_T^A\overline{Y}_T^B\mid\mathbf{X}\right]-\mathbb{E}\left[\overline{Y}_T^A\mid\mathbf{X}\right]\mathbb{E}\left[\overline{Y}_T^B\mid\mathbf{X}\right]\\&=\beta(\beta+\tau)+\frac{1}{PJ}\mathbb{E}\left[v_i^2\mid\mathbf{X}\right]\\&\quad+\frac{1}{(PJ)^2mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathbb{E}[\omega_{it}\omega_{is}\mid\mathbf{X}]-\beta(\beta+\tau)\\&=\frac{\sigma_v^2}{PJ}+\frac{1}{(PJ)^2mr}\sum_{i=1}^{PJ}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\omega_{it},\omega_{is}\mid\mathbf{X})\\&=\frac{1}{PJ}\left(\sigma_v^2+\psi_{i,T}^X\right)\end{aligned} \tag{B.36}$$

Similarly, with arbitrary correlations, the remaining covariance terms of Equation (B.12) are

$$\mathrm{Cov}\left(\overline{Y}_T^A,\overline{Y}_C^A\mid\mathbf{X}\right)=\frac{1}{r}\left(\sigma_\delta^2+\psi_{t,X}^A\right) \tag{B.37}$$

$$\mathrm{Cov}\left(\overline{Y}_T^B,\overline{Y}_C^B\mid\mathbf{X}\right)=\frac{1}{m}\left(\sigma_\delta^2+\psi_{t,X}^B\right) \tag{B.38}$$

$$\mathrm{Cov}\left(\overline{Y}_C^A,\overline{Y}_C^B\mid\mathbf{X}\right)=\frac{1}{(1-P)J}\left(\sigma_v^2+\psi_{i,C}^X\right) \tag{B.39}$$

Substituting Equations (B.32)–(B.39) into Equation (B.12) and simplifying gives

$$\mathrm{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)=\frac{1}{PJr}\left(r\sigma_v^2+PJ\sigma_\delta^2+\sigma_\omega^2+(r-1)\psi_{i,T}^A+(PJ-1)\psi_{t,T}^A\right)$$

$$
\begin{aligned}
&+ \frac{1}{PJm}\left(m\sigma_{\upsilon}^2 + PJ\sigma_{\delta}^2 + \sigma_{\omega}^2 + (m-1)\psi_{i,T}^B + (PJ-1)\psi_{t,T}^B\right) \\
&+ \frac{1}{(1-P)Jr}\left(r\sigma_{\upsilon}^2 + (1-P)J\sigma_{\delta}^2 + \sigma_{\omega}^2 + (r-1)\psi_{i,C}^A + ((1-P)J-1)\psi_{t,C}^A\right) \\
&+ \frac{1}{(1-P)Jm}\left(m\sigma_{\upsilon}^2 + (1-P)J\sigma_{\delta}^2 + \sigma_{\omega}^2 + (m-1)\psi_{i,C}^B + ((1-P)J-1)\psi_{t,C}^B\right) \\
&- 2\left[\frac{1}{PJ}\left(\sigma_{\upsilon}^2 + \psi_{i,T}^X\right) + \frac{1}{r}\left(\sigma_{\delta}^2 + \psi_{t,X}^A\right)\right. \\
&\qquad \left. + \frac{1}{m}\left(\sigma_{\delta}^2 + \psi_{t,X}^B\right) + \frac{1}{(1-P)J}\left(\sigma_{\upsilon}^2 + \psi_{i,C}^X\right)\right] \\
&= \left(\frac{m+r}{P(1-P)Jmr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{PJm}\right)\psi_{i,T}^B + \left(\frac{PJ-1}{PJm}\right)\psi_{t,T}^B \\
&+ \left(\frac{r-1}{PJr}\right)\psi_{i,T}^A + \left(\frac{PJ-1}{PJr}\right)\psi_{t,T}^A + \left(\frac{m-1}{(1-P)Jm}\right)\psi_{i,C}^B \\
&+ \left(\frac{(1-P)J-1}{(1-P)Jm}\right)\psi_{t,C}^B + \left(\frac{r-1}{(1-P)Jr}\right)\psi_{i,C}^A + \left(\frac{(1-P)J-1}{(1-P)Jr}\right)\psi_{t,C}^A \\
&- \frac{2}{PJ}\psi_{i,T}^X - \frac{2}{m}\psi_{t,X}^B - \frac{2}{r}\psi_{t,X}^A - \frac{2}{(1-P)J}\psi_{i,C}^X
\end{aligned}
$$

Minimum detectable effect With arbitrary correlations, the MDE is

$$
\begin{aligned}
MDE &= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)} \\
&= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\left[\left(\frac{m+r}{P(1-P)Jmr}\right)\sigma_{\omega}^2 + \left(\frac{m-1}{PJm}\right)\psi_{i,T}^B + \left(\frac{PJ-1}{PJm}\right)\psi_{t,T}^B\right. \\
&\qquad + \left(\frac{r-1}{PJr}\right)\psi_{i,T}^A + \left(\frac{PJ-1}{PJr}\right)\psi_{t,T}^A + \left(\frac{m-1}{(1-P)Jm}\right)\psi_{i,C}^B \\
&\qquad + \left(\frac{(1-P)J-1}{(1-P)Jm}\right)\psi_{t,C}^B + \left(\frac{r-1}{(1-P)Jr}\right)\psi_{i,C}^A \\
&\qquad + \left(\frac{(1-P)J-1}{(1-P)Jr}\right)\psi_{t,C}^A - \frac{2}{PJ}\psi_{i,T}^X - \frac{2}{m}\psi_{t,X}^B - \frac{2}{r}\psi_{t,X}^A \\
&\qquad \left. - \frac{2}{(1-P)J}\psi_{i,C}^X\right]^{1/2}
\end{aligned} \tag{B.40}
$$

Lemma 1 in Appendix B.1.3 shows that Equation (B.40) is equal in expectation to the SCR power calculation formula when treatment is randomly assigned.

B.1.2.4 Collapsed dataset

Bertrand, Duflo, and Mullainathan (2004, henceforth BDM) suggest an alternative to the CRVE in order to achieve the correct false rejection rates in the presence of serial correlation: ignore the time-series structure of the data by averaging the pre-treatment data and the post-treatment data for each unit, then estimate a panel DD regression on this two-period

collapsed dataset and apply the OLS variance estimator. While this does yield the desired false rejection rate, simply applying the McKenzie formula to a collapsed dataset will *not* yield the desired power.

Using the model from Appendix B.1.2.2 (under Assumptions B6–B10), suppose that prior to estimation, the data are collapsed to 1 pre- and 1 post-treatment observation per unit (to eliminate serial correlation, as suggested in BDM). The resulting DGP for the collapsed dataset is

$$Y_{ip}^{C} = \beta + \tau D_{ip}^{C} + \upsilon_i + \delta_p^{C} + \omega_{ip}^{C} \tag{B.41}$$

where Y_{ip}^{C} is the average outcome for unit i for collapsed period p and the other variables are as defined in Appendix B.1.2.2 or are the collapsed analogs. Note that the τ in Equation (B.41) is equivalent to that in Equation (B.7). These models will yield the same estimate of the treatment effect, $\widehat{\tau}$, but different estimates of its variance in the presence of a (pre-collapsed) serially correlated error structure.

Equivalence of first-difference model BDM show that applying the OLS variance estimator to Equation (B.41) achieves the correct false rejection rate. To see why this is the case, note that this collapsed model can alternatively be expressed as a first-difference model by subtracting each unit's collapsed pre-treatment data from its collapsed post-treatment data. Let ΔY_i^{C} be this difference for the outcome of interest, which gives

$$\begin{aligned}
\Delta Y_i^{C} &= Y_{iA}^{C} - Y_{iB}^{C} \\
&= \left(\beta + \tau D_{iA}^{C} + \upsilon_i + \delta_A^{C} + \omega_{iA}^{C}\right) - \left(\beta + \tau D_{iB}^{C} + \upsilon_i + \delta_B^{C} + \omega_{iB}^{C}\right) \\
&= \tau\left(D_{iA}^{C} - D_{iB}^{C}\right) + \left(\delta_A^{C} - \delta_B^{C}\right) + \left(\omega_{iA}^{C} - \omega_{iB}^{C}\right)
\end{aligned}$$

Defining the other differences variables similarly gives the first-difference DGP of

$$\Delta Y_i^{C} = \tau \Delta D_i^{C} + \Delta\delta^{C} + \Delta\omega_i^{C} \tag{B.42}$$

Equations (B.41) and (B.42) are isomorphic, so estimating these models yields not only the same estimate of the treatment effect, $\widehat{\tau}$, but also the same estimate of its variance. Note that the first-difference model, Equation (B.42), is cross-sectional, so the error terms are i.i.d. and the model meets the assumptions of OLS.[2] As a result, the OLS variance estimator is unbiased for the first-difference model, as well as the isomorphic collapsed model of Equation (B.41).

Power Although using a collapsed dataset yields the correct false rejection (or Type I error) rate, experiments will not be correctly powered if the McKenzie formula is applied to a collapsed dataset. To see this, first consider the MDE of an experiment based on the first-difference model of Equation (B.42). This is a cross-sectional model, so applying Equations (B.5) and (B.6) yields:

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \frac{\sigma^2_{\Delta\omega^C}}{P(1-P)J} \tag{B.43}$$

[2]There are no cross-sectional error correlations due to Assumption B9, because randomization obviates the need to account for this kind of correlation (see Lemma 1).

$$MDE = \left(t_{1-\kappa}^{J-2} + t_{\alpha/2}^{J-2}\right)\sqrt{\frac{\sigma_{\Delta\omega^C}^2}{P(1-P)J}}$$

where $\sigma_{\Delta\omega^C}^2$ is the variance of the error term in the collapsed, first-difference model. This variance can be expressed as a function of the parameters that define the error structure of the *collapsed* data

$$\begin{aligned}\sigma_{\Delta\omega^C}^2 &= \text{Var}\left(\omega_{iA}^C - \omega_{iB}^C \mid \mathbf{X}\right) \\ &= \text{Var}\left(\omega_{iA}^C \mid \mathbf{X}\right) + \text{Var}\left(\omega_{iB}^C \mid \mathbf{X}\right) - 2\,\text{Cov}\left(\omega_{iA}^C, \omega_{iB}^C \mid \mathbf{X}\right) \\ &= 2\left(\sigma_{\omega^C}^2 - \psi_{\omega^C}^X\right) \qquad \text{(B.44)}\end{aligned}$$

where $\sigma_{\omega^C}^2$ is the variance of the error term of the collapsed model, and $\psi_{\omega^C}^X$ is the average covariance between error terms for the same unit in the collapsed model. Substituting Equation (B.44) into Equation (B.43) gives the variance of $\hat{\tau}$ in terms of parameters of the collapsed data:

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \left(\frac{2}{P(1-P)J}\right)\left(\sigma_{\omega^C}^2 - \psi_{\omega^C}^X\right) \qquad \text{(B.45)}$$

This formula is equal to the variance of $\hat{\tau}$ from the SCR formula applied to collapsed data, where $m = r = 1$. Applying the McKenzie formula to collapsed data, however, gives the incorrect variance:

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \left(\frac{2}{P(1-P)J}\right)\sigma_{\omega^C}^2 \qquad \text{(B.46)}$$

Equations (B.45) and (B.46) differ by the $\psi_{\omega^C}^X$ term that characterizes the covariance between the pre- and post-treatment error terms in the collapsed data. This term is omitted from the McKenzie formula that (incorrectly) assumes no serial correlation in the error structure. The error structure parameters of the collapsed data in Equation (B.45) can also be expressed as functions of the parameters that define the error structure of the original, uncollapsed data

$$\begin{aligned}\sigma_{\omega^C}^2 &= \frac{1}{2}\text{Var}\left(\omega_{iA}^C \mid \mathbf{X}\right) + \frac{1}{2}\text{Var}\left(\omega_{iB}^C \mid \mathbf{X}\right) \\ &= \frac{1}{2}\text{Var}\left(\frac{1}{r}\sum_{t=1}^{r}\omega_{it} \mid \mathbf{X}\right) + \frac{1}{2}\text{Var}\left(\frac{1}{m}\sum_{s=-m+1}^{0}\omega_{is} \mid \mathbf{X}\right) \\ &= \frac{1}{2r}\left[\sigma_\omega^2 + (r-1)\psi^A\right] + \frac{1}{2m}\left[\sigma_\omega^2 + (m-1)\psi^B\right] \\ &= \frac{1}{2}\left[\left(\frac{m+r}{mr}\right)\sigma_\omega^2 + \left(\frac{r-1}{r}\right)\psi^A + \left(\frac{m-1}{m}\right)\psi^B\right] \qquad \text{(B.47)} \\ \psi_{\omega^C}^X &= \frac{1}{J}\sum_{i=1}^{J}\text{Cov}\left(\omega_{iA}^C, \omega_{iB}^C \mid \mathbf{X}\right) \\ &= \frac{1}{J}\sum_{i=1}^{J}\text{Cov}\left(\frac{1}{r}\sum_{t=1}^{r}\omega_{it}, \frac{1}{m}\sum_{s=-m+1}^{0}\omega_{is} \mid \mathbf{X}\right)\end{aligned}$$

$$= \frac{1}{Jmr} \sum_{i=1}^{J} \sum_{t=1}^{r} \sum_{s=-m+1}^{0} \text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X})$$

$$= \psi^X \tag{B.48}$$

Substituting Equations (B.47) and (B.48) into Equation (B.45) gives

$$\text{Var}\left(\hat{\tau} \mid \mathbf{X}\right) = \left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_\omega^2 + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]$$

which is equivalent to the variance of $\hat{\tau}$ given in Equation (B.31) that is a component of the SCR formula. Hence, the uncollapsed, collapsed, and first-difference models yield (virtually) equivalent MDEs when using the appropriate power calculation formula.[3] By contrast, the McKenzie formula ignores the between-period serial correlation that remains after collapsing serially correlated data.

B.1.2.5 Analysis of covariance (ANCOVA)

Model There are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption B15 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a time-invariant unit effect distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption B16 (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*, and* $J - 1$ *unit dummies. This again follows from random assignment of* D_{it}*.*

Assumption B17 (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

[3]The only difference between these three *MDE* calculations is the critical values $t^d_{1-\kappa}$ and $t^d_{\alpha/2}$. The uncollapsed model will apply the CRVE *ex post*, which implies $d = J$ degrees of freedom *ex ante*. The collapsed model will apply the OLS variance estimator *ex post* to a panel with $2J$ observations and $J + 2$ regressors, which implies $d = J - 2$ degrees of freedom *ex ante*. The first-difference model will apply the OLS variance estimator *ex post* to a cross-sectional specification with J observations and 2 regressors, which implies $d = J - 2$ degrees of freedom *ex ante*.

Assumption B18 (Independence across units)**.** $\mathbb{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ \forall\ t, s.$

Assumption B19 (Uniform covariance structures)**.** *Define:*

$$\psi_i^B \equiv \frac{2}{m(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi_i^A \equiv \frac{2}{r(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi_i^X \equiv \frac{1}{mr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of unit i, respectively. Using these definitions, assume that $\psi^B = \psi_i^B$, $\psi^A = \psi_i^A$, and $\psi^X = \psi_i^X\ \forall i$.

In this section, we consider an analysis of covariance (ANCOVA) regression model:

$$Y_{it} = \alpha + \tau D_i + \theta \overline{Y}_i^B + \varepsilon_{it} \tag{B.49}$$

where Y_{it}, τ, and D_i are defined as above; α is an intercept term, $\overline{Y}_i^B = \frac{1}{m}\sum_{t=-m+1}^{0} Y_{it}$ is the mean pre-period value of the outcome variable for unit i, and ε_{it} is an idiosyncratic error term. Note that this regression is estimated using only post-treatment outcomes as the dependent variable, with a unit's pre-treatment mean outcome entering as an independent variable.

Note that this following derivations necessitate two relatively strong assumptions, as we discuss in Section 3.2.2.2 of the main text. First, we assume zero time shocks in the data generating process, thereby omitting δ_t from Assumption B15. If we included time shocks as in Assumption B3, then the conditional expectations in Equations (B.56) and (B.57) would each depend on the error terms and pre-period means of all units in the experiment, making the analytic solution intractable.[4] Second, Assumption B19 forces the ψ parameters to be constant across all units. Without this simplifying assumption, Equations (B.58)–(B.60) would contain many additional terms describing the heterogeneity of ψ parameters, which

[4]If we include time shocks and time fixed effects in our data generating process and regression model, respectively, the simplest analytic solution includes de-meaning by time period to remove these time shocks, similar to Appendix B.1.2.1. Then Equations (B.56) and (B.57) would instead describe the expectation of the product of the de-meaned error terms conditional on the de-meaned pre-period averages, which are a function of the error terms and pre-period means of all units in the experiment. These expressions increase in complexity in proportion to the number of units, making them analytically intractable for any reasonable number of experimental units while allowing for arbitrary serial correlation. An alternate solution method does not de-mean by time period to remove time shocks, but instead solves directly for all parameters in the model. This method requires inverting a $(m + r + 2) \times (m + r + 2)$ matrix, which is analytically intractable for an experiment of arbitrary length.

has a negligible effect in experiments of resaonable size[5] and cannot be estimated from a pre-existing dataset.[6]

Coefficient estimate The coefficient estimates from an OLS regression are

$$\begin{aligned}
\begin{pmatrix}\hat{\alpha}\\ \hat{\tau}\\ \hat{\theta}\end{pmatrix} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \left(\begin{pmatrix}\mathbf{1}'\\ \mathbf{D}'\\ \overline{\mathbf{Y}}^{\mathbf{B}\prime}\end{pmatrix}\begin{pmatrix}\mathbf{1} & \mathbf{D} & \overline{\mathbf{Y}}^{\mathbf{B}}\end{pmatrix}\right)^{-1}\begin{pmatrix}\mathbf{1}'\\ \mathbf{D}'\\ \overline{\mathbf{Y}}^{\mathbf{B}\prime}\end{pmatrix}\mathbf{Y} \\
&= \underbrace{\begin{pmatrix} Jr & PJr & Jr\overline{Y}^{B} \\ PJr & PJr & PJr\overline{Y}^{B}_{T} \\ Jr\overline{Y}^{B} & PJr\overline{Y}^{B}_{T} & r\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2\end{pmatrix}}_{\mathbf{M}}{}^{-1}\begin{pmatrix} Jr\overline{Y}^{A} \\ PJr\overline{Y}^{A}_{T} \\ r\sum_{i=1}^{J}\overline{Y}^{A}_{i}\overline{Y}^{B}_{i}\end{pmatrix} \qquad \text{(B.50)}
\end{aligned}$$

The inverse of the $\mathbf{M}$ matrix is

$$\mathbf{M}^{-1} = \frac{PJ^2r^2}{\det(\mathbf{M})}\times\begin{pmatrix} \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - P(\overline{Y}^{B}_{T})^2 & \overline{Y}^{B}\overline{Y}^{B}_{T} - \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 & P\overline{Y}^{B}_{T} - \overline{Y}^{B} \\ \overline{Y}^{B}\overline{Y}^{B}_{T} - \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 & \frac{1}{PJ}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - \frac{1}{P}(\overline{Y}^{B})^2 & \overline{Y}^{B} - \overline{Y}^{B}_{T} \\ P\overline{Y}^{B}_{T} - \overline{Y}^{B} & \overline{Y}^{B} - \overline{Y}^{B}_{T} & 1-P\end{pmatrix}$$

and the determinant of $\mathbf{M}$ is

$$\begin{aligned}
\det(\mathbf{M}) &= PJ^3r^3\left[\frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - P(\overline{Y}^{B}_{T})^2\right] + P^2J^3r^3\left[\overline{Y}^{B}\overline{Y}^{B}_{T} - \frac{1}{J}\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2\right] \\
&\quad + PJ^3r^3\overline{Y}^{B}\left[P\overline{Y}^{B}_{T} - \overline{Y}^{B}\right] \\
&= P(1-P)J^2r^3\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - P^2(1-P)(\overline{Y}^{B}_{T})^2 - P(1-P)^2(\overline{Y}^{B}_{C})^2 \\
&= P(1-P)J^2r^3\left[\sum_{i=1}^{PJ}(\overline{Y}^{B}_{i} - \overline{Y}^{B}_{T})^2 + \sum_{i=PJ+1}^{J}(\overline{Y}^{B}_{i} - \overline{Y}^{B}_{C})^2\right]
\end{aligned}$$

Therefore, the coefficient estimates are

$$\hat{\alpha} = \frac{\overline{Y}^{A}_{C}\left[\sum_{i=1}^{J}(\overline{Y}^{B}_{i})^2 - PJ(\overline{Y}^{B}_{T})^2 - (1-P)J(\overline{Y}^{B}_{C})^2\right]}{\sum_{i=1}^{J}(\overline{Y}^{B}_{i} - \overline{Y}^{B}_{T})^2 + \sum_{i=PJ+1}^{J}(\overline{Y}^{B}_{i} - \overline{Y}^{B}_{C})^2}$$

[5]We test this second assumption by simulating datasets with uniform ψ parameters or heterogeneous ψ parameters and then performing Monte Carlo simulations as described in Appendix B.2.1. We find no appreciable difference in the simulated power of these different datasets.

[6]A dataset contains only one realiztion of each error draw, so it is impossible to estimate a covariance between any two of these draws. We can, however, estimate an average covariance over all units by treating each unit's error as a draw from the same distribution, as described in more detail in Appendices B.4.1 and B.5.1.

$$
\begin{aligned}
&\quad - \frac{\overline{Y}_C^B \left[\sum_{i=1}^J \overline{Y}_i^B \overline{Y}_i^A - PJ\overline{Y}_T^B \overline{Y}_T^A - (1-P)J\overline{Y}_C^B \overline{Y}_T^A\right]}{\sum_{i=1}^J (\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^J (\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
\hat{\tau} &= \frac{(\overline{Y}_T^A - \overline{Y}_C^A)\left[\sum_{i=1}^J (\overline{Y}_i^B)^2 - PJ(\overline{Y}_T^B)^2 - (1-P)J(\overline{Y}_C^B)^2\right]}{\sum_{i=1}^J (\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^J (\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
&\quad - \frac{(\overline{Y}_T^B - \overline{Y}_C^B)\left[\sum_{i=1}^J \overline{Y}_i^B \overline{Y}_i^A - PJ\overline{Y}_T^B \overline{Y}_T^A - (1-P)J\overline{Y}_C^B \overline{Y}_T^A\right]}{\sum_{i=1}^J (\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^J (\overline{Y}_i^B - \overline{Y}_C^B)^2} \\
\hat{\theta} &= \frac{\sum_{i=1}^J \overline{Y}_i^B \overline{Y}_i^A - PJ\overline{Y}_T^B \overline{Y}_T^A - (1-P)J\overline{Y}_C^B \overline{Y}_C^A}{\sum_{i=1}^J (\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^J (\overline{Y}_i^B - \overline{Y}_C^B)^2}
\end{aligned}
$$

We rewrite these terms as

$$
\begin{aligned}
\hat{\alpha} &= \overline{Y}_C^A - \hat{\theta}\overline{Y}_C \\
\hat{\tau} &= (\overline{Y}_T^A - \overline{Y}_C^A) - \hat{\theta}(\overline{Y}_T^B - \overline{Y}_C^B) \\
\hat{\theta} &= \frac{\sum_{i=1}^{PJ} (\overline{Y}_i^B - \overline{Y}_T^B)(\overline{Y}_i^A - \overline{Y}_T^A) + \sum_{i=PJ+1}^J (\overline{Y}_i^B - \overline{Y}_C^B)(\overline{Y}_i^A - \overline{Y}_C^A)}{\sum_{i=1}^{PJ} (\overline{Y}_i^B - \overline{Y}_T^B)^2 + \sum_{i=PJ+1}^J (\overline{Y}_i^B - \overline{Y}_C^B)^2}
\end{aligned}
$$

Variance of coefficient estimates The variance of each coefficient estimate is:

$$
\text{Var}\left[\begin{pmatrix}\hat{\alpha}\\ \hat{\tau}\\ \hat{\theta}\end{pmatrix} \mid \mathbf{X}\right] = \underbrace{(\mathbf{X}'\mathbf{X})}_{\mathbf{M}}^{-1} \underbrace{\mathbf{X}'\,\mathbb{E}[\varepsilon\varepsilon' \mid \mathbf{X}]\mathbf{X}}_{\mathbf{N}} \underbrace{(\mathbf{X}'\mathbf{X})}_{\mathbf{M}}^{-1}
$$

where $\mathbf{M}$ is the same as in Equation (B.50) and $\mathbf{N}$ is

$$
\begin{aligned}
\mathbf{N} &= \begin{pmatrix}\mathbf{1}'\\ \mathbf{D}'\\ \overline{\mathbf{Y}}^{\mathbf{B}\prime}\end{pmatrix} \begin{pmatrix}\mathbb{E}[\varepsilon_{11}\varepsilon_{11} \mid \mathbf{X}] & \dots & \mathbb{E}[\varepsilon_{11}\varepsilon_{Jr} \mid \mathbf{X}]\\ \vdots & \ddots & \vdots \\ \mathbb{E}[\varepsilon_{Jr}\varepsilon_{11} \mid \mathbf{X}] & \dots & \mathbb{E}[\varepsilon_{Jr}\varepsilon_{Jr} \mid \mathbf{X}]\end{pmatrix} \begin{pmatrix}\mathbf{1} & \mathbf{D} & \overline{\mathbf{Y}}^{\mathbf{B}}\end{pmatrix} \\
&= \sum_{t=1}^r \sum_{s=1}^r \begin{pmatrix} \sum_{i=1}^J \sum_{j=1}^J \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ} \sum_{j=1}^J \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^J \sum_{j=1}^J \overline{Y}_i^B\, \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] \\ \sum_{i=1}^{PJ} \sum_{j=1}^J \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ} \sum_{j=1}^{PJ} \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ} \sum_{j=1}^J \overline{Y}_j^B\, \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] \\ \sum_{i=1}^J \sum_{j=1}^J \overline{Y}_i^B\, \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^{PJ} \sum_{j=1}^{PJ} \overline{Y}_j^B\, \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] & \sum_{i=1}^J \sum_{j=1}^J \overline{Y}_i^B \overline{Y}_j^B\, \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] \end{pmatrix}
\end{aligned}
$$

We now focus on the variance of the treatment effect estimator, $\hat{\tau}$. Combining this expression for $\mathbf{N}$ with the expression for $\mathbf{M}^{-1}$ above, this variance is given by

$$
\begin{aligned}
\text{(B.51)}\quad \text{Var}(\hat{\tau} \mid \mathbf{X}) = \frac{1}{J^2 r^2 Z^2} \sum_{t=1}^r \sum_{s=1}^r \Bigg\{ &\frac{1}{P^2} \sum_{i=1}^{PJ} \sum_{j=1}^{PJ} \Big[\Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B)\Big) \\
&\times \Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_j^B - \overline{Y}_T^B)\Big)\, \mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}]\Big] \\
&+ \frac{2}{P(1-P)} \sum_{i=1}^{PJ} \sum_{j=PJ+1}^{J} \Big[\Big(Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B)\Big)
\end{aligned}
$$

$$\times \left(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_j^B-\overline{Y}_C^B)\right)\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid \mathbf{X}]\Bigg]$$
$$+\frac{1}{(1-P)^2}\sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\Bigg[\left(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_C^B)\right)$$
$$\times \left(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_j^B-\overline{Y}_C^B)\right)\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid \mathbf{X}]\Bigg]\Bigg\}$$

where

$$Z=\sum_{k=1}^{PJ}(\overline{Y}_k^B-\overline{Y}_T^B)^2+\sum_{k=PJ+1}^{J}(\overline{Y}_k^B-\overline{Y}_C^B)^2$$

We consider each pair of summations in Equation (B.51) by defining

$$V_T\equiv\frac{1}{P^2}\sum_{t=1}^{r}\sum_{s=1}^{r}\sum_{i=1}^{PJ}\sum_{j=1}^{PJ}\Bigg[\left(Z-PJ(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_T^B)\right)$$
$$\times\left(Z-PJ(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_j^B-\overline{Y}_T^B)\right)\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}]\Bigg] \tag{B.52}$$

$$V_X\equiv\frac{2}{P(1-P)}\sum_{t=1}^{r}\sum_{s=1}^{r}\sum_{i=1}^{PJ}\sum_{j=PJ+1}^{J}\Bigg[\left(Z-PJ(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_T^B)\right)$$
$$\times\left(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_j^B-\overline{Y}_C^B)\right)\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}]\Bigg] \tag{B.53}$$

$$V_C\equiv\frac{1}{(1-P)^2}\sum_{t=1}^{r}\sum_{s=1}^{r}\sum_{i=PJ+1}^{J}\sum_{j=PJ+1}^{J}\Bigg[\left(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_C^B)\right)$$
$$\times\left(-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_j^B-\overline{Y}_C^B)\right)\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}]\Bigg] \tag{B.54}$$

so Equation (B.51) can be rewritten as

$$\mathrm{Var}(\hat{\tau}\mid\mathbf{X})=\frac{1}{J^2r^2Z^2}\left(V_T+V_X+V_C\right) \tag{B.55}$$

Each of these expressions contains an expectation term, which we rewrite, using the Law of Iterated Expectations, as

$$\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid\mathbf{X}]=\mathbb{E}[\varepsilon_{it}\varepsilon_{js}\mid\overline{Y}_i^B,\overline{Y}_j^B]=\mathbb{E}\left[\varepsilon_{js}\,\mathbb{E}[\varepsilon_{it}\mid\varepsilon_{js},\overline{Y}_i^B,\overline{Y}_j^B]\mid\overline{Y}_i^B,\overline{Y}_j^B\right]$$

Note that ε_{it}, ε_{js}, $\overline{Y}_i^B$, and $\overline{Y}_j^B$ are random variables drawn from a multivariate normal distribution. Consider two cases. First, if $i\neq j$, then the inner expectation term is

$$\mathbb{E}[\varepsilon_{it}\mid\varepsilon_{js},\overline{Y}_i^B,\overline{Y}_j^B]=\mathbb{E}[\varepsilon_{it}\mid\overline{Y}_i^B]=\frac{\mathrm{Cov}(\varepsilon_{it},\overline{Y}_i^B)}{\mathrm{Var}(\overline{Y}_i^B)}\overline{Y}_i^B$$

and the full expectation term is

$$\begin{aligned}\mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] &= \frac{\operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B)}{\operatorname{Var}(\overline{Y}_i^B)}\overline{Y}_i^B \,\mathbb{E}[\varepsilon_{js} \mid \overline{Y}_j^B] \\ &= \frac{\operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B)}{\operatorname{Var}(\overline{Y}_i^B)}\overline{Y}_i^B \frac{\operatorname{Cov}(\varepsilon_{js}, \overline{Y}_j^B)}{\operatorname{Var}(\overline{Y}_j^B)}\overline{Y}_j^B \end{aligned} \tag{B.56}$$

However, note that

$$\operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B) = \operatorname{Cov}(\omega_{it}, \overline{\omega}_i^B) - r\psi^X$$

so summing over all r post-treatent periods gives

$$\sum_{t=1}^{r} \operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B) = \frac{1}{m}\sum_{t=1}^{r}\sum_{s=-m+1}^{0} \operatorname{Cov}(\omega_{it}, \omega_{is}) - r\psi^X = 0$$

As a result, the case of $i \neq j$ will not contribute to V_T, V_X, and V_C. Consider instead the case of $i = j$, in which case the inner expectation term is

$$\begin{aligned}\mathbb{E}[\varepsilon_{it} \mid \varepsilon_{js}, \overline{Y}_i^B, \overline{Y}_j^B] &= \mathbb{E}[\varepsilon_{it} \mid \varepsilon_{is}, \overline{Y}_i^B] \\ &= \frac{\operatorname{Var}(\overline{Y}_i^B)\operatorname{Cov}(\varepsilon_{it}, \varepsilon_{is}) - \operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B)\operatorname{Cov}(\varepsilon_{is}, \overline{Y}_i^B)}{\operatorname{Var}(\varepsilon_{is})\operatorname{Var}(\overline{Y}_i^B) - \operatorname{Cov}(\varepsilon_{is}, \overline{Y}_i^B)^2}\varepsilon_{is} \\ &\quad + \frac{\operatorname{Var}(\varepsilon_{is})\operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B) - \operatorname{Cov}(\varepsilon_{it}, \varepsilon_{is})\operatorname{Cov}(\varepsilon_{is}, \overline{Y}_i^B)}{\operatorname{Var}(\varepsilon_{is})\operatorname{Var}(\overline{Y}_i^B) - \operatorname{Cov}(\varepsilon_{is}, \overline{Y}_i^B)^2}\overline{Y}_i^B\end{aligned}$$

and the full expectation term is

$$\begin{aligned}\mathbb{E}[\varepsilon_{it}\varepsilon_{js} \mid \mathbf{X}] &= \mathbb{E}\left[\varepsilon_{is}\,\mathbb{E}[\varepsilon_{it} \mid \varepsilon_{is}, \overline{Y}_i^B] \mid \overline{Y}_i^B\right] \\ &= \operatorname{Cov}(\varepsilon_{it}, \varepsilon_{is}) - \frac{\operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B)\operatorname{Cov}(\varepsilon_{is}, \overline{Y}_i^B)}{\operatorname{Var}(\overline{Y}_i^B)} \\ &\quad + \frac{\operatorname{Cov}(\varepsilon_{it}, \overline{Y}_i^B)\operatorname{Cov}(\varepsilon_{is}, \overline{Y}_i^B)}{\operatorname{Var}(\overline{Y}_i^B)^2}(\overline{Y}_i^B)^2\end{aligned} \tag{B.57}$$

Substituting Equations (B.56) and (B.57) into Equations (B.52)–(B.54) and simplifying gives

$$\begin{aligned}V_T = \frac{r^2}{P^2}&\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \\ &\times \sum_{i=1}^{PJ}\left[Z - PJ(\overline{Y}_T^B - \overline{Y}_C^B)(\overline{Y}_i^B - \overline{Y}_T^B))\right]^2\end{aligned} \tag{B.58}$$

$$V_X = 0 \tag{B.59}$$

$$V_C = \frac{r^2}{(1-P)^2}\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \times \sum_{i=PJ+1}^{J}\left[-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_C^B))\right]^2 \tag{B.60}$$

where θ is the same as in Equation (B.49) and can be expressed as

$$\theta = \frac{m\sigma_v^2 + m\psi^X}{m\sigma_v^2 + \sigma_\omega^2 + (m-1)\psi^B}$$

Then substituting Equations (B.58)–(B.60) into Equation (B.55) gives the variance of the treatment effect estimator:

$$\begin{aligned}
\mathrm{Var}(\hat{\tau} \mid \mathbf{X}) &= \frac{1}{J^2r^2Z^2}\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \\
&\quad \times \left[\frac{r^2}{P^2}\sum_{i=1}^{PJ}\left[Z - PJ(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_T^B))\right]^2\right. \\
&\quad \left. + \frac{r^2}{(1-P)^2}\sum_{i=PJ+1}^{J}\left[-Z-(1-P)J(\overline{Y}_T^B-\overline{Y}_C^B)(\overline{Y}_i^B-\overline{Y}_C^B))\right]^2\right] \\
&= \frac{1}{J^2Z^2}\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right] \\
&\quad \times \left[\frac{JZ^2}{P(1-P)} + J^2(\overline{Y}_T^B-\overline{Y}_C^B)^2Z\right] \\
&= \left[\frac{1}{P(1-P)J} + \frac{(\overline{Y}_T^B-\overline{Y}_C^B)^2}{Z}\right] \\
&\quad \times \left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]
\end{aligned} \tag{B.61}$$

The second component of the first term in Equation (B.61), $\frac{(\overline{Y}_T^B-\overline{Y}_C^B)^2}{Z}$, is the only component of this formula that includes data, rather than parameters describing the experimental design or error structure. Additionally, this term decreases in expectation to zero as the number of units in the experiment increases, so it is negligible when J is sufficiently large. Accordingly, we drop this term to yield an approximate expression for the variance of the treatment effect estimator:

$$\mathrm{Var}(\hat{\tau} \mid \mathbf{X}) \approx \frac{1}{P(1-P)J}\left[(1-\theta)^2\sigma_v^2 + \left(\frac{\theta^2}{m}+\frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]$$

Minimum detectable effect When using an ANCOVA regression model, the MDE is

$$MDE = \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right)\sqrt{\mathrm{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)}$$

$$= \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right) \left[\frac{1}{P(1-P)J} + \frac{(\overline{Y}_T^B - \overline{Y}_C^B)^2}{Z}\right]^{1/2}$$
$$\times \left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]^{1/2}$$
$$\approx \left(t_{1-\kappa}^J + t_{\alpha/2}^J\right) \tag{B.62}$$
$$\times \sqrt{\frac{1}{P(1-P)J}\left[(1-\theta)^2\sigma_\upsilon^2 + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma_\omega^2 + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]}$$

This is the serial-correlation-robust ANCOVA (SCR ANCOVA) power calculation formula, found in Equation (3.11) in the main text.

B.1.3 Arbitrary cross-sectional correlations

In this section, we provide proofs of Lemma 1 from the main text. We begin by proving the analogous cross-sectional case, in Lemma B1:[7]

Lemma B1. *In a cross-sectional model with random assignment to treatment,* $\frac{\sigma_\varepsilon^2}{P(1-P)J}$ *is an unbiased estimator of the expectation of* $\text{Var}(\hat{\tau} \mid \mathbf{X})$ *even if* $\mathbb{E}[\varepsilon_i\varepsilon_j \mid \mathbf{X}] \neq 0$ *for some* $i \neq j$.

Proof We wish to demonstrate that the OLS variance estimator, which assumes that errors are independent across units, is an unbiased estimator of the *ex ante* expected variance under non-i.i.d. errors when units are randomly assigned to treatment. To do this, consider a model similar to that of Appendix B.1.1 but allowing for correlation in error terms. That is:

Model There are J units randomly assigned a treatment status D_i, with proportion P in treatment ($D_i = 1$) and proportion $(1-P)$ in control ($D_i = 0$). The units are indexed so $i \in [1, PJ]$ is treated and $j \in [PJ+1, J]$ is a control. We make standard assumptions for randomized trials:

Assumption B20 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_i = \beta + \tau D_i + \varepsilon_i$$

where Y_i *is the outcome of interest for unit* i*;* β *is the expected outcome of non-treated units;* τ *is the treatment effect which is homogeneous across all units;* D_i *is a treatment indicator; and* ε_i *is an idiosyncratic error term distributed (not necessarily i.i.d.)* $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Assumption B21 (Strict exogeneity)**.** $\mathbb{E}[\varepsilon_i \mid \mathbf{X}] = 0$*, where* $\mathbf{X} = [\mathbf{1}\ \mathbf{D}]$*. In practice, this follows from random assignment of* D_i.

[7] Campbell (1977) provides the first version of this proof, which is cited by Moulton (1986), and which imposes a grouped error structure. Our proof allows for arbitrary cross-sectional error dependence. Athey and Imbens (2017b) and Athey and Imbens (2017a) still recommend using Eicker-Huber-White standard errors in this case, to allow for heteroskedasticity. To our knowledge, no paper discusses power calculations in the presence of heteroskedastic disturbances.

Coefficient estimate The coefficient estimates from an OLS regression are the same as in Appendix B.1.1:

$$\begin{aligned}\widehat{\beta} &= \overline{Y}_C \\ \widehat{\tau} &= \overline{Y}_T - \overline{Y}_C\end{aligned}$$

Variance of coefficient estimate The variance of the estimate of the treatment effect, $\hat{\tau}$, is

$$\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) = \text{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) + \text{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) - 2\,\text{Cov}\left(\overline{Y}_T, \overline{Y}_C \mid \mathbf{X}\right) \tag{B.63}$$

where the first term of Equation (B.63) is

$$\begin{aligned}\text{Var}\left(\overline{Y}_T \mid \mathbf{X}\right) &= \text{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} Y_i \mid \mathbf{X}\right) \\ &= \text{Var}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} \varepsilon_i \mid \mathbf{X}\right) \\ &= \frac{\sigma_\varepsilon^2}{PJ} + \frac{2}{(PJ)^2}\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ} \text{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X}) \\ &= \frac{\sigma_\varepsilon^2}{PJ} + \frac{PJ-1}{PJ}\psi_T \qquad \text{(B.64)}\end{aligned}$$

where

$$\psi_T \equiv \frac{2}{PJ(PJ-1)}\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ} \text{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)$$

is the average covariance between treated units. Similarly, the second term of Equation (B.63) is

$$\text{Var}\left(\overline{Y}_C \mid \mathbf{X}\right) = \frac{\sigma_\varepsilon^2}{(1-P)J} + \frac{(1-P)J-1}{(1-P)J}\psi_C \tag{B.65}$$

where

$$\psi_C \equiv \frac{2}{(1-P)J((1-P)J-1)}\sum_{i=PJ+1}^{J-1}\sum_{j=i+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)$$

is the average covariance between control units. The third term of Equation (B.63) is

$$-2\,\text{Cov}\left(\overline{Y}_T, \overline{Y}_C \mid \mathbf{X}\right) = -2\,\text{Cov}\left(\frac{1}{PJ}\sum_{i=1}^{PJ} Y_i, \frac{1}{(1-P)J}\sum_{j=PJ+1}^{J} Y_j \mid \mathbf{X}\right)$$

$$= \frac{-2}{P(1-P)J^2} \sum_{i=1}^{PJ} \sum_{j=PJ+1}^{J} \text{Cov}(\varepsilon_i, \varepsilon_j \mid \mathbf{X})$$

$$= -2\psi_{TC} \tag{B.66}$$

where

$$\psi_{TC} \equiv \frac{1}{P(1-P)J^2} \sum_{i=1}^{PJ} \sum_{j=PJ+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)$$

is the average covariance between treatment and control units. Substituting Equations (B.64)–(B.66) into Equation (B.63) yields

$$\begin{aligned}\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right) &= \left[\frac{\sigma_\varepsilon^2}{PJ} + \frac{PJ-1}{PJ}\psi_T\right] + \left[\frac{\sigma_\varepsilon^2}{(1-P)J} + \frac{(1-P)J-1}{(1-P)J}\psi_C\right] - 2\psi_{TC} \\ &= \frac{\sigma_\varepsilon^2}{P(1-P)J} + \frac{PJ-1}{PJ}\psi_T + \frac{(1-P)J-1}{(1-P)J}\psi_C - 2\psi_{TC}\end{aligned}$$

Note that σ_ε^2, P, and J are constant population or design parameters. With this in mind, taking expectations yields

$$\mathbb{E}\left[\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = \frac{\sigma_\varepsilon^2}{P(1-P)J} + \frac{PJ-1}{PJ}\,\mathbb{E}\left[\psi_T\right] + \frac{(1-P)J-1}{(1-P)J}\,\mathbb{E}\left[\psi_C\right] - 2\,\mathbb{E}\left[\psi_{TC}\right] \tag{B.67}$$

By random assignment to treatment, the expectation terms are

$$\begin{aligned}\mathbb{E}\left[\psi_T\right] &= \frac{2}{PJ(PJ-1)}\,\mathbb{E}\left[\sum_{i=1}^{PJ-1}\sum_{j=i+1}^{PJ} \text{Cov}\left(\varepsilon_i, \varepsilon_j \mid \mathbf{X}\right)\right] \\ &= \frac{2}{PJ(PJ-1)} \sum_{i=1}^{J-1}\sum_{j=i+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j\right)\mathbb{E}\left[\mathbf{1}\{i \in T, j \in T\}\right] \\ &= \frac{2P}{J(PJ-1)} \sum_{i=1}^{J-1}\sum_{j=i+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j\right) \qquad \text{(B.68)}\end{aligned}$$

$$\mathbb{E}\left[\psi_C\right] = \frac{2(1-P)}{J((1-P)J-1)} \sum_{i=1}^{J-1}\sum_{j=i+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j\right) \tag{B.69}$$

$$\mathbb{E}\left[\psi_{TC}\right] = \frac{2}{J^2} \sum_{i=1}^{J-1}\sum_{j=i+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j\right) \tag{B.70}$$

Substituting Equations (B.68)–(B.70) into Equation (B.67) yields

$$\mathbb{E}\left[\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = \frac{\sigma_\varepsilon^2}{P(1-P)J} + \frac{2}{J^2} \sum_{i=1}^{J-1}\sum_{j=i+1}^{J} \text{Cov}\left(\varepsilon_i, \varepsilon_j\right)$$

$$+\frac{2}{J^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\text{Cov}\left(\varepsilon_i,\varepsilon_j\right)-\frac{4}{J^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\text{Cov}\left(\varepsilon_i,\varepsilon_j\right)$$

$$=\frac{\sigma_\varepsilon^2}{P(1-P)J} \tag{B.71}$$

By Assumptions B20 and B21, the OLS variance estimator is an unbiased estimator of $\frac{\sigma_\varepsilon^2}{P(1-P)J}$. That is:

$$\mathbb{E}\left[\widehat{\text{Var}}_{OLS}\left(\widehat{\tau}\mid\mathbf{X}\right)\right]=\frac{\sigma_\varepsilon^2}{P(1-P)J} \tag{B.72}$$

Combining Equations (B.71) and (B.72) gives

$$\mathbb{E}\left[\text{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)\right]=\mathbb{E}\left[\widehat{\text{Var}}_{OLS}\left(\widehat{\tau}\mid\mathbf{X}\right)\right] \tag{B.73}$$

Therefore, the OLS variance estimator is an unbiased estimator of the *ex ante* expected variance under random assignment to treatment, even under non-i.i.d. errors. □

Lemma 1. *In a panel difference-in-differences model with treatment randomly assigned at the unit level,* $\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma_\omega^2+\left(\frac{m-1}{m}\right)\psi^B+\left(\frac{r-1}{r}\right)\psi^A-2\psi^X\right]$ *is an unbiased estimator of the expectation of* $\text{Var}(\hat{\tau}\mid\mathbf{X})$*, even in the presence of arbitrary within-period cross-sectional correlations.*

Proof We wish to demonstrate that the serial-correlation-robust variance, which assumes that errors are independent across units, is an unbiased estimator of the *ex ante* expected variance under arbitrary within-period correlations, when units are randomly assigned to treatment. To do this, again consider the model of Appendix B.1.2.3.

Coefficient estimate The coefficient estimate is the same as in Appendix B.1.2.3:

$$\widehat{\tau}=\left(\overline{Y}_T^A-\overline{Y}_T^B\right)-\left(\overline{Y}_C^A-\overline{Y}_C^B\right)$$

Variance of coefficient estimate The variance of the treatment effect estimator, $\widehat{\tau}$, is also the same as in Appendix B.1.2.3:

$$\begin{aligned}\text{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)=&\left(\frac{m+r}{P(1-P)Jmr}\right)\sigma_\omega^2+\left(\frac{m-1}{PJm}\right)\psi_{i,T}^B+\left(\frac{PJ-1}{PJm}\right)\psi_{t,T}^B\\&+\left(\frac{r-1}{PJr}\right)\psi_{i,T}^A+\left(\frac{PJ-1}{PJr}\right)\psi_{t,T}^A+\left(\frac{m-1}{(1-P)Jm}\right)\psi_{i,C}^B\\&+\left(\frac{(1-P)J-1}{(1-P)Jm}\right)\psi_{t,C}^B+\left(\frac{r-1}{(1-P)Jr}\right)\psi_{i,C}^A+\left(\frac{(1-P)J-1}{(1-P)Jr}\right)\psi_{t,C}^A\\&-\frac{2}{PJ}\psi_{i,T}^X-\frac{2}{m}\psi_{t,X}^B-\frac{2}{r}\psi_{t,X}^A-\frac{2}{(1-P)J}\psi_{i,C}^X\end{aligned}$$

Next, we show that, in expectation, this is equal to the variance component of the SCR formula. We begin by taking expectations:

$$\begin{aligned}(\text{B.74})\quad \mathbb{E}\left[\operatorname{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = &\left(\frac{m+r}{P(1-P)Jmr}\right)\sigma_\omega^2 + \left(\frac{m-1}{PJm}\right)\mathbb{E}\left[\psi_{i,T}^B\right] + \left(\frac{PJ-1}{PJm}\right)\mathbb{E}\left[\psi_{t,T}^B\right] \\ &+ \left(\frac{r-1}{PJr}\right)\mathbb{E}\left[\psi_{i,T}^A\right] + \left(\frac{PJ-1}{PJr}\right)\mathbb{E}\left[\psi_{t,T}^A\right] \\ &+ \left(\frac{m-1}{(1-P)Jm}\right)\mathbb{E}\left[\psi_{i,C}^B\right] + \left(\frac{(1-P)J-1}{(1-P)Jm}\right)\mathbb{E}\left[\psi_{t,C}^B\right] \\ &+ \left(\frac{r-1}{(1-P)Jr}\right)\mathbb{E}\left[\psi_{i,C}^A\right] + \left(\frac{(1-P)J-1}{(1-P)Jr}\right)\mathbb{E}\left[\psi_{t,C}^A\right] \\ &- \frac{2}{PJ}\mathbb{E}\left[\psi_{i,T}^X\right] - \frac{2}{m}\mathbb{E}\left[\psi_{t,X}^B\right] - \frac{2}{r}\mathbb{E}\left[\psi_{t,X}^A\right] - \frac{2}{(1-P)J}\mathbb{E}\left[\psi_{i,C}^X\right]\end{aligned}$$

where

$$\mathbb{E}\left[\psi_{i,T}^B\right] = \frac{2}{Jm(m-1)}\sum_{i=1}^{J}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\operatorname{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathbb{E}\left[\psi_{t,T}^B\right] = \frac{2P}{J(PJ-1)m}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\operatorname{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathbb{E}\left[\psi_{i,T}^A\right] = \frac{2}{Jr(r-1)}\sum_{i=1}^{J}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathbb{E}\left[\psi_{t,T}^A\right] = \frac{2P}{J(PJ-1)r}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\operatorname{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathbb{E}\left[\psi_{i,C}^B\right] = \frac{2}{Jm(m-1)}\sum_{i=1}^{J}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\operatorname{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathbb{E}\left[\psi_{t,C}^B\right] = \frac{2(1-P)}{J((1-P)J-1)m}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\operatorname{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathbb{E}\left[\psi_{i,C}^A\right] = \frac{2}{Jr(r-1)}\sum_{i=1}^{J}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathbb{E}\left[\psi_{t,C}^A\right] = \frac{2(1-P)}{J((1-P)J-1)r}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\operatorname{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathbb{E}\left[\psi_{i,T}^X\right] = \frac{1}{Jmr}\sum_{i=1}^{J}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\operatorname{Cov}\left(\omega_{it},\omega_{is}\right)$$

$$\mathbb{E}\left[\psi_{t,TC}^B\right] = \frac{2}{J^2m}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\operatorname{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathbb{E}\left[\psi^A_{t,TC}\right] = \frac{2}{J^2 r}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right)$$

$$\mathbb{E}\left[\psi^X_{i,C}\right] = \frac{1}{Jmr}\sum_{i=1}^{J}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{is}\right)$$

Using these terms, we rewrite Equation (B.74) as

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Var}\left(\widehat{\tau}\mid\mathbf{X}\right)\right] &= \left(\frac{m+r}{P(1-P)Jmr}\right)\sigma^2_\omega + \left(\frac{m-1}{PJm}\right)\psi^B \\
&\quad + \frac{2}{J^2m^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) + \left(\frac{r-1}{PJr}\right)\psi^A \\
&\quad + \frac{2}{J^2r^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) + \left(\frac{m-1}{(1-P)Jm}\right)\psi^B \\
&\quad + \frac{2}{J^2m^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) + \left(\frac{r-1}{(1-P)Jr}\right)\psi^A \\
&\quad + \frac{2}{J^2r^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) - \left(\frac{2}{PJ}\right)\psi^X \\
&\quad - \frac{4}{J^2m^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) - \left(\frac{2}{(1-P)J}\right)\psi^X \\
&\quad - \frac{4}{J^2r^2}\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&= \left(\frac{m+r}{P(1-P)Jmr}\right)\sigma^2_\omega + \left[\frac{m-1}{PJm} + \frac{m-1}{(1-P)Jm}\right]\psi^B \\
&\quad + \left[\frac{r-1}{PJr} + \frac{r-1}{(1-P)Jr}\right]\psi^A - \left[\frac{2}{PJ} + \frac{2}{(1-P)J}\right]\psi^X \\
&\quad + \left[\frac{2}{J^2m^2} + \frac{2}{J^2m^2} - \frac{4}{J^2m^2}\right]\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=-m+1}^{0}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&\quad + \left[\frac{2}{J^2r^2} + \frac{2}{J^2r^2} - \frac{4}{J^2r^2}\right]\sum_{i=1}^{J-1}\sum_{j=i+1}^{J}\sum_{t=1}^{r}\mathrm{Cov}\left(\omega_{it},\omega_{jt}\right) \\
&= \frac{1}{P(1-P)J}\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right] \qquad \text{(B.75)}
\end{aligned}$$

Recall from Equation (B.31) that the variance component of the SCR formula is

$$\mathrm{Var}_{SCR}\left(\widehat{\tau}\mid\mathbf{X}\right) = \frac{1}{P(1-P)J}\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right] \qquad \text{(B.76)}$$

Combining Equations (B.75) and (B.76) gives

$$\mathbb{E}\left[\text{Var}\left(\widehat{\tau} \mid \mathbf{X}\right)\right] = \text{Var}_{SCR}\left(\widehat{\tau} \mid \mathbf{X}\right)$$

Therefore, the SCR variance estimator is an unbiased estimator of the *ex ante* expected variance under random assignment to treatment, even under non-i.i.d. errors. □

B.1.4 Equivalence of alternative difference-in-difference estimators

In this section, we demonstrate that the DD treatment effect estimator—and therefore, the variance of this estimator—with a full set of unit and time fixed effects is the same as the "simplified" DD estimator where unit fixed effects are replaced with a treatment group dummy and time fixed effects are replaced with a post-period dummy. We begin with the same model, data generating process, and assumptions as Section B.1.2.1, suppressed here in the interest of parsimony.

Coefficient estimate In Section B.1.2.1 above, we estimate the treatment effect, τ, using OLS with unit and time fixed effects. Here, we instead estimate the τ using OLS with a treatment group dummy and a post-treatment period dummy (i.e. Equation (3.8), with new notation):

$$Y_{it} = \beta + \tau D_{it} + \alpha_g + \gamma_p + \varepsilon_{it} \tag{B.77}$$

where α_g is a treatment group dummy variable and γ_p is a post-treatment-period dummy variable. In a balanced panel, this is equivalent to de-meaning by the group and post-treatment period. Define:

$$\ddot{Y}_{it} = Y_{it} - \overline{Y}_g - \overline{Y}_p + \overline{\overline{Y}} \tag{B.78}$$

$$\ddot{D}_{it} = D_{it} - \overline{D}_g - \overline{D}_p + \overline{\overline{D}} \tag{B.79}$$

$$\ddot{\varepsilon}_{it} = \varepsilon_{it} - \overline{\varepsilon}_g - \overline{\varepsilon}_p + \overline{\overline{\varepsilon}} \tag{B.80}$$

where:

$$\overline{Y}_T = \frac{1}{PJ(m+r)} \sum_{i=1}^{PJ} \sum_{t=-m+1}^{r} Y_{it}$$

$$\overline{Y}_C = \frac{1}{(1-P)J(m+r)} \sum_{i=PJ+1}^{J} \sum_{t=-m+1}^{r} Y_{it}$$

$$\overline{Y}_A = \frac{1}{Jr} \sum_{i=1}^{J} \sum_{t=1}^{r} Y_{it}$$

$$\overline{Y}_B = \frac{1}{Jm} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} Y_{it}$$

$$\overline{\overline{Y}} = \frac{1}{J(m+r)} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} Y_{it}$$

with $\overline{D}_g$, $\overline{D}_p$, $\overline{\overline{D}}$, $\overline{\varepsilon}_g$, $\overline{\varepsilon}_p$, and $\overline{\overline{\varepsilon}}$ defined analogously. Substituting Equations (B.78)–(B.80) into Equation (B.77) and simplifying gives the de-meaned DGP:

$$\begin{aligned}
\ddot{Y}_{it} &= (\tau D_{it} + \alpha_g + \gamma_p + \varepsilon_{it}) - (\tau \overline{D}_g + \alpha_g + \overline{\gamma} + \overline{\varepsilon}_g) \\
&\quad - (\tau \overline{D}_p + \overline{\alpha} + \gamma_p + \overline{\varepsilon}_p) + (\tau \overline{\overline{D}} + \overline{\alpha} + \overline{\gamma} + \overline{\overline{\varepsilon}}) \\
&= \tau(D_{it} - \overline{D}_g - \overline{D}_p + \overline{\overline{D}}) + (\varepsilon_{it} - \overline{\varepsilon}_g - \overline{\varepsilon}_p + \overline{\overline{\varepsilon}}) \\
&= \tau \ddot{D}_{it} + \ddot{\varepsilon}_{it}
\end{aligned}$$

The estimate of the treatment effect is

$$\begin{aligned}
\widehat{\tau} &= (\ddot{\mathbf{D}}'\ddot{\mathbf{D}})^{-1}\ddot{\mathbf{D}}'\ddot{\mathbf{Y}} \\
&= \left(\sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it}^2 \right)^{-1} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} \ddot{Y}_{it} \\
&= \frac{m+r}{P(1-P)Jmr} \left[\sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} Y_{it} - \sum_{i=1}^{J} \overline{Y}_i \sum_{t=-m+1}^{r} \ddot{D}_{it} \right. \\
&\qquad \left. - \sum_{t=-m+1}^{r} \overline{Y}_t \sum_{i=1}^{J} \ddot{D}_{it} + \overline{Y} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} \right] \\
&= \frac{m+r}{P(1-P)Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{r} \ddot{D}_{it} Y_{it} \\
&= \frac{m+r}{P(1-P)Jmr} \left[\sum_{i=1}^{PJ} \left(\frac{(1-P)m}{m+r} \sum_{t=1}^{r} Y_{it} - \frac{(1-P)r}{m+r} \sum_{t=-m+1}^{0} Y_{it} \right) \right. \\
&\qquad \left. + \sum_{i=PJ+1}^{J} \left(\frac{Pr}{m+r} \sum_{t=-m+1}^{0} Y_{it} - \frac{Pm}{m+r} \sum_{t=1}^{r} Y_{it} \right) \right] \\
&= \left(\overline{Y}_T^A - \overline{Y}_T^B \right) - \left(\overline{Y}_C^A - \overline{Y}_C^B \right)
\end{aligned}$$

where

$$\begin{aligned}
\overline{Y}_T^A &= \frac{1}{PJr} \sum_{i=1}^{PJ} \sum_{t=1}^{r} Y_{it} \\
\overline{Y}_T^B &= \frac{1}{PJm} \sum_{i=1}^{PJ} \sum_{t=-m+1}^{0} Y_{it} \\
\overline{Y}_C^A &= \frac{1}{(1-P)Jr} \sum_{i=PJ+1}^{J} \sum_{t=1}^{r} Y_{it}
\end{aligned}$$

$$\overline{Y}_C^B = \frac{1}{(1-P)Jm} \sum_{i=PJ+1}^{J} \sum_{t=-m+1}^{0} Y_{it}$$

which is identical to Equation (B.11), meaning that the variance is also identical. □

It is trivial to extend the same logic to variants on the DD estimator in which *either* unit fixed effects are replaced with a treatment group dummy or time fixed effects are replaced with a post-treatment dummy.

B.2 Figures in main text

This section provides further detail on the simulations and power calculations referenced in the main text. We discuss the algorithms and assumptions behind each of the simulation plots, as well as the two analytical power calculation figures.

B.2.1 Simulated AR(1) data

In Figure 3.1, we run Monte Carlo simulations where each iteration generates a new simulated dataset with an idiosyncratic error term (ω_{it}) that evolves via an AR(1) process. We vary the following three parameters across sets of 10,000 simulations: the number of pre-treatment periods (m), the number of post-treatment periods (r), and the AR(1) dependence parameter (γ). The remaining parameters (J, P, α, κ, β, μ_υ, σ^2_υ, μ_δ, σ^2_δ, σ^2_ω) are fixed across all simulations.

Step 1: We calculate τ^{McK} and τ^{SCR} for each set of simulations, given a set of parameters values for m, r, and γ. These values are functions of the number of pre-treatment periods m, the number of post-treatment periods r, the number of units J, the proportion of units randomized into treatment P, the desired Type-I error rate α, the desired power κ, the idiosyncratic variance σ^2_ω, and (for τ^{SCR}) the error structure as defined by ψ^B, ψ^A, and ψ^X:

$$\tau^{McK} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \sqrt{\left(\frac{\sigma^2_\omega}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)}$$

$$\tau^{SCR} = (t^J_{1-\kappa} + t^J_{\alpha/2}) \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}$$

Note that for both τ^{McK} and τ^{SCR}, we calculate the critical values $t^J_{1-\kappa}$ and $t^J_{\alpha/2}$ assuming J degrees of freedom, which is consistent with applying the CRVE *ex post*, clustering at the unit level with J units. Note also that ψ^B, ψ^A, and ψ^X depend on the correlation structure of the errors, and the AR(1) process enables us to derive closed form expressions for these covariances in terms of γ, σ^2_ω, m, and r. Because we set $m = r$ across all simulations, we can write ψ^B, ψ^A, and ψ^X as:

$$\psi^B = \frac{2\sigma^2_\omega}{(m-1)m} \sum_{z=1}^{m-1} (m-z)\gamma^z \tag{B.81}$$

$$\psi^A = \psi^B \tag{B.82}$$

$$\psi^X = \frac{\sigma_\omega^2}{m^2}\left[\sum_{z=1}^{m} z\gamma^z + \sum_{z=m+1}^{2m-1}(2m-z)\gamma^z\right] \tag{B.83}$$

Step 2: For each simulation, we generate a dataset as specified by the data generating process:

$$Y_{it} = \beta + \upsilon_i + \delta_t + \omega_{it}$$

To do this, we draw J independent values of υ_i from the distribution $N(\mu_\upsilon, \sigma_\upsilon^2)$, and draw $m + r$ independent values of δ_t from the distribution $N(\mu_\delta, \sigma_\delta^2)$. We create the idiosyncratic error $\omega_{it} = \gamma\omega_{i(t-1)} + \xi_{it}$ by simulating an AR(1) process with serial correlation γ and a white noise term ξ_{it} drawn from the distribution $N(0, \sigma_\xi^2)$, where $\sigma_\xi^2 = \sigma_\omega^2(1-\gamma^2)$.[8]

Step 3: We randomly assign treatment to PJ units. This involves randomly scrambling a vector of PJ ones and $(1-P)J$ zeros and assigning each unit i either a 1 indicating treatment or a 0 indicating control.[9] This allows us to construct a time-varying treatment indicator D_{it}, where $D_{it} = 1$ for all treated units in post-treatment periods only and $D_{it} = 0$ otherwise. We then create three outcome variables by adding treatment effects to the data generated in the previous step:

$$\begin{aligned} Y_{it}^{McK} &\equiv Y_{it} + \tau^{McK} D_{it} \\ Y_{it}^{SCR} &\equiv Y_{it} + \tau^{SCR} D_{it} \\ Y_{it}^{0} &\equiv Y_{it} + \tau^{0} D_{it}, \end{aligned}$$

where $\tau^0 = 0$ is a placebo treatment effect.

Step 4: We separately estimate the following three OLS-fixed effects regressions:

$$\begin{aligned} Y_{it}^{McK} &= \beta + \tau^{McK} D_{it} + \upsilon_i + \delta_t + \omega_{it} \\ Y_{it}^{SCR} &= \beta + \tau^{SCR} D_{it} + \upsilon_i + \delta_t + \omega_{it} \\ Y_{it}^{0} &= \beta + \tau^{0} D_{it} + \upsilon_i + \delta_t + \omega_{it} \end{aligned}$$

Step 5: For each estimated $\hat{\tau}^{McK}$, $\hat{\tau}^{SCR}$, and $\hat{\tau}^0$, we compute both OLS standard errors and CRVE standard errors, clustered at the unit level.

We repeat Steps 2–5 10,000 times, for values of $m = r \in \{1, \ldots, 20\}$ and for values of $\gamma \in \{0, 0.3, 0.5, 0.7, 0.9\}$.[10] After each set of 10,000 simulations, we calculate the percent of simulations where $\hat{\tau}^{McK}$, $\hat{\tau}^{SCR}$, and $\hat{\tau}^0$ reject the null hypothesis of $\tau = 0$ at significance level α, under both OLS and CRVE standard errors.

[8] We allow for a sufficiently long "burn-in period" in this AR(1) process, so that the process starts to evolve many periods before the first period of simulated data.

[9] Our code rounds PJ to the nearest integer value, even though PJ is already an integer in our main parameterization. Note that for the τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR} to be precisely calibrated, the effective $\tilde{P}$ (where $\tilde{P} = \text{round}(PJ)/J$) needs to equal the actual parameter value P.

[10] We set $m = r$ only for simplicity. However, the results are very similar if we fix $m = 3$ and vary $r \in \{1, \ldots, 20\}$, or vice versa. In Appendix B.3.1, we present results that vary m and r separately.

Figure 3.1 reports these rejection rates on the vertical axes, with the number of pre- and post-treatment periods ($m = r$) on the horizontal axes, for each value of γ. Reading the top row left to right, we report the rejection rates for τ^{McK} under OLS standard errors, for τ^{McK} under CRVE standard errors, and for τ^{SCR} under CRVE standard errors, respectively. Because these are rejection rates of true effects, we interpret these curves as realized statistical power. Reading the bottom row left to right, we report the rejection rates for τ^0 under OLS standard errors, for τ^0 under CRVE standard errors, and for τ^0 under CRVE standard errors, respectively. Because these are rejection rates of placebo effects, we report these curves as realized false rejection rates. (The bottom-center and bottom-right panels report identical rejection rates, because the center and right columns have the same test for false rejection rates.)

We fix $\alpha = 0.05$ and $\kappa = 0.80$ across all simulations, as these are the critical values commonly used in practice. However, they are essentially arbitrary, and our simulation results would look identical if we had chosen alternative tolerances for Type I vs. Type II errors. (The only difference would be that the vertical axes would change to reflect these alternative values.) All other fixed parameter values are arbitrary. We have set $J = 500$, $P = 0.5$, $\beta = 1$, $\mu_\upsilon = 100$, $\sigma^2_\upsilon = 80$, $\mu_\delta = 20$, $\sigma^2_\delta = 10$, and $\sigma^2_\omega = 10$. These values of σ^2_υ, σ^2_δ, and σ^2_ω imply an intracluster correlation coefficient of $\rho_\upsilon = 0.8$ and within-period correlation coefficient of $\rho_\delta = 0.1$. Importantly, our simulation results do not depend on any particular combination of these parameters values, because they rely on the internal consistency of the *ex ante* treatment effect calibration and the *ex post* estimation, conditional on a *given* set of parameter values. The only exceptions are for J and P: J must be larger enough to allow us to use the CRVE estimator (i.e., at least 40 clusters), and P must be within a reasonable range (i.e., between 0.1 and 0.9) such that there are a sufficient number of both treated clusters and control clusters.

We report additional simulations for DD power calculations with AR(1) data, each of which slightly tweaks the above algorithm. Figure 3.2 uses $m \in \{1, 2, 3, 4, 5, 6\}$ and $r \in \{1, 2, 3, 4, 5, 6\}$, varying the number of pre-treatment and post-treatment period separately. Figure 3.3 alters Step 4 above, by either (i) replacing υ_i fixed effects with a Treat_i dummy, (ii) replacing δ_t fixed effects with a Post_t dummy, or (iii) both. Figure 3.5 combines this tweak to Step 4 with alternate parameterization of the DGP in Step 2: (i) $\sigma^2_\upsilon = 0$, (ii) $\sigma^2_\delta = 0$, or (iii) $\sigma^2_\upsilon = \sigma^2_\delta = 0$.

Importantly, removing unit fixed effects alters the asymptotic properties of the CRVE, as the number of clusters no longer increases one-for-one with the number of regressors. For all simulations in this paper and appendix, we apply the CRVE estimator that is appropriate to each estimating equation (following Cameron and Miller (2015), pp. 14–15): if the regression includes unit fixed effects and we are clustering by unit, we scale standard errors by $\sqrt{\frac{N-1}{N-J-1}}$, where N the number of observations in the regression and J is the number of units (clusters); if the regression does not include unit fixed effects, we do not apply this correction factor.

The ANCOVA simulations in Figure 3.4 follow the a very similar algorithm as those in Figure 3.1. Each of the above steps is identical, except for the following changes:

Step 1: Instead of using the DD power calculation formulas, we apply the analogous ANCOVA formulas to calculate τ^{McK} and τ^{SCR}:[11]

$$\tau^{McK} = (t^J_{1-k} + t^J_{\alpha/2})\sqrt{\frac{1}{P(1-P)J}\left[(1-\breve{\theta})^2\sigma^2_\upsilon + \left(\frac{\breve{\theta}^2}{m} + \frac{1}{r}\right)\sigma^2_\omega\right]}$$

where $\breve{\theta} = \dfrac{m\sigma^2_\upsilon}{m\sigma^2_\upsilon + \sigma^2_\omega}$, and

$$\tau^{SCR} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\left[\frac{1}{P(1-P)J} + \frac{(\overline{Y}^B_T - \overline{Y}^B_C)^2}{Z}\right]^{1/2} \times \left[(1-\theta)^2\sigma^2_\upsilon + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma^2_\omega + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X\right]^{1/2}$$

where

$$\theta = \frac{m\sigma^2_\upsilon + m\psi^X}{m\sigma^2_\upsilon + \sigma^2_\omega + (m-1)\psi^B}$$
$$Z = \sum_{i=1}^{PJ}(\overline{Y}^B_i - \overline{Y}^B_T)^2 + \sum_{i=PJ+1}^{J}(\overline{Y}^B_i - \overline{Y}^B_C)^2$$

Step 2: We set $\sigma^2_\delta = 0$, consistent with the assumptions of our analytic ANCOVA derivations. This assumption would be unrealistic for real-world data, however this Monte Carlo exercise focuses on the internal (in)consistencies of the ANCOVA power calculation formulas.

Step 3: We also collapse all pre-treatment observations into unit-specific unweighted averages:

$$\overline{Y}^B_i \equiv \sum_{t=-m+1}^{0} Y_{it}$$

Step 4: We separately estimate the following ANCOVA regressions:

$$Y^{McK}_{it} = \beta + \tau^{McK}D_i + \theta\overline{Y}^B_i + \varepsilon_{it}$$
$$Y^{SCR}_{it} = \beta + \tau^{SCR}D_i + \theta\overline{Y}^B_i + \varepsilon_{it}$$
$$Y^{0}_{it} = \beta + \tau^{0}D_i + \theta\overline{Y}^B_i + \varepsilon_{it}$$

Note that we do not include time fixed effects δ_t, as we have assumed away time shocks for these simulations.

[11] To create Figure 3.4, we use the exact SCR ANCOVA formula of Equation (B.62) rather than the approximate formula of Equation (3.11).

Step 5: We compute only CRVE standard errors, clustered at the unit level.

Figure 3.4 reports the resulting rejection rates of the ANCOVA estimator. We do not report the false rejection rates from the τ^0 regressions for the sake of brevity, and they do achieve the desired $\alpha = 0.05$ rejection rate in all cases.

An additional nuance with the ANCOVA simulations is that our simulation results are now sensitive to the intracluster correlation coefficient ρ_υ. This is because the proportion of variance that is unit-specific now affects the precision of the $\hat{\tau}$ estimator, because we have replaced the unit fixed effect (which directly controlled for this variance) with a linear control in the average pre-treatment level of the outcome variable. Figure 3.4 sets $\rho_\upsilon = 0.8$, and the results are similar for alternative values of ρ_υ.

B.2.2 Bloom et al. (2015) data

In Figure 3.6, we run Monte Carlo simulations using data from Bloom et al. (2015). These simulations are analogous to those described above, except that rather than simulating data, we use an actual dataset from a published panel RCT. We downloaded the paper's dataset from the *Quarterly Journal of Economics* website, and focused on the data used to estimate the paper's main results, reported in Column (1) of Table II of the paper. Consistent with the regression model that produced this result, we base our Monte Carlo analysis on the following DD specification:

$$\text{Performance}_{it} = \alpha \text{Treat}_i \times \text{Experiment}_t + \beta_t + \gamma_i + \varepsilon_{it}$$

Converting the original paper's notation to our notation, and substituting the outcome variable and fixed effects with the names of the variables in the Bloom et al. (2015) dataset, we have:

$$\underbrace{\texttt{perform1}_{it}}_{Y_{it}} = \tau D_{it} + \underbrace{\texttt{year_week}_t}_{\delta_t} + \underbrace{\texttt{personid}_i}_{\upsilon_i} + \omega_{it}$$

We keep only units in the main sample (i.e. $\texttt{expgroup} \in \{0, 1\}$), only pre-treatment weeks of data (i.e. $\texttt{year_week} < 201049$), and only individuals with non-missing $\texttt{perform1}_{it}$ values for all weeks of pre-treatment data. This leaves us with a balanced panel of $J = 79$ individuals across 48 weeks. Table B.1 provides summary statistics of the resulting dataset.

We conduct simulations on this dataset by varying the number of pre-treatment periods (m) and the number of post-treatment periods (r). As with the simulations described above, we vary the panel length for values $m = r \in \{1, \ldots, 20\}$, iterating 10,000 simulations for each value of $m = r$. We set the parameter values $\sigma^2_\omega = 0.507$ and $\gamma = 0.233$ by regressing Y_{it} on person and week fixed effects, calculating the variance of the resulting residuals $\hat{\omega}_{it}$, and then estimating $\hat{\gamma}$ using

$$\hat{\omega}_{it} = \gamma \hat{\omega}_{i(t-1)} + \xi_{it}$$

Step 1: We calculate τ^{McK} and $\tau^{AR(1)}$ for each set of simulations, given $m = r$:

$$\tau^{McK} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \sqrt{\left(\frac{\sigma^2_\omega}{P(1-P)J}\right)\left(\frac{m+r}{mr}\right)}$$

Table B.1: Summary statistics – Bloom et al. (2015)

Mean	Std. Dev.	Min	Max	AR(1) $\hat{\gamma}$	Individuals	Periods	Observations
0.153	0.943	−2.766	3.665	0.233	79	48	3,792

Notes: This table shows summary statistics for worker productivity in the Bloom et al. (2015) data. The data are weekly job performance z-scores, constructed by taking the average of normalized performance measures, where each measure is standardized to have a mean of 0 and standard deviation of 1 across the sample. Our sample consists only of individuals that had no missing observations throughout the entire pre-treatment period, January 1, 2010 through November 28, 2010. We compute $\hat{\gamma}$ by estimating Equation (3.6) on residuals from this dataset. In doing so, we cluster standard errors at the individual level. The 95% confidence interval is [0.165, 0.300]. For more details on the standardized job performance measures and the actual experimental design, see Bloom et al. (2015).

$$\tau^{AR(1)} = (t^J_{1-\kappa} + t^J_{\alpha/2}) \times \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\breve{\psi}^B + \left(\frac{r-1}{r}\right)\breve{\psi}^A - 2\breve{\psi}^X\right]}$$

where

$$\breve{\psi}^B = \frac{2\sigma^2_\omega}{(m-1)m}\sum_{z=1}^{m-1}(m-z)\gamma^z$$
$$\breve{\psi}^A = \breve{\psi}^B$$
$$\breve{\psi}^X = \frac{\sigma^2_\omega}{m^2}\left[\sum_{z=1}^{m} z\gamma^z + \sum_{z=m+1}^{2m-1}(2m-z)\gamma^z\right]$$

We denote the covariance terms as $\breve{\psi}^B$, $\breve{\psi}^A$, and $\breve{\psi}^X$ to indicate that the AR(1) error assumption is a (poor) representation of the more complex covariance structure of this dataset. For both τ^{McK} and $\tau^{AR(1)}$, we calculate the critical values $t^J_{1-\kappa}$ and $t^J_{\alpha/2}$ assuming J degrees of freedom, which is consistent with applying the CRVE *ex post*, clustering at the individual level with J individuals.

Step 2: We calculate τ^{SCR} given $m = r$, by non-parametrically estimating $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from residuals. Appendix B.4.1 provides step-by-step details of this estimation algorithm. Rather than impose an AR(1) structure on the serial correlation, this method enables us to flexibly characterize the covariance structure of the Bloom et al. (2015) dataset with just three averaged parameters. This allows us to calculate τ^{SCR} as:

$$\tau^{SCR} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \times \sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)k_\sigma\sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right)k_B\psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right)k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}}\right]}$$

where

$$k_\sigma = \frac{I(m+r)^2}{2(I-1)mr}$$
$$k_B = \frac{I(m+r)^2}{2(I-1)r^2}$$
$$k_A = \frac{I(m+r)^2}{2(I-1)m^2}$$
$$k_X = 0$$

Appendix B.5 provides a derivation of the coefficients k_σ, k_B, k_A, and k_X, and it proves that this expression for τ^{SCR} as a function of *estimated* variance-covariance parameters is equal (in expectation) to the MDE as a function of the *true* variance-covariance parameters.[12]

Step 3: For each simulation, we randomly select a range of $m + r$ consecutive weeks in the dataset. This subset of weeks will become the $(m + r)$-period panel dataset used in this particular simulation. We randomly assign treatment to PJ individuals. This involves randomly scrambling a vector of PJ ones and $(1 - P)J$ zeros and assigning each individual i either a 1 indicating treatment or a 0 indicating control.[13] This allows us to construct a time-varying treatment indicator D_{it}, where $D_{it} = 1$ for all treated units in post-treatment periods only and $D_{it} = 0$ otherwise. We then create three outcome variables by adding treatment effects to the data generated in the previous step:

$$Y_{it}^{McK} \equiv Y_{it} + \tau^{McK} D_{it}$$
$$Y_{it}^{AR(1)} \equiv Y_{it} + \tau^{AR(1)} D_{it}$$
$$Y_{it}^{SCR} \equiv Y_{it} + \tau^{SCR} D_{it}$$

Step 4: We separately estimate the following three OLS-fixed effects regressions:

$$Y_{it}^{McK} = \beta + \tau^{McK} D_{it} + \upsilon_i + \delta_t + \omega_{it}$$
$$Y_{it}^{AR(1)} = \beta + \tau^{AR(1)} D_{it} + \upsilon_i + \delta_t + \omega_{it}$$
$$Y_{it}^{SCR} = \beta + \tau^{SCR} D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

Step 5: For each estimated $\hat{\tau}^{McK}$, $\hat{\tau}^{AR(1)}$, and $\hat{\tau}^{SCR}$, we compute CRVE standard errors, clustered at the individual level.

As with the AR(1) simulations above, we repeat Steps 3–5 10,000 times, for values of $m = r \in \{1, \ldots, 12\}$.[14] After each set of 10,000 simulations, we calculate the percent of

[12] I denotes the number of units used to estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$ and $\psi^X_{\hat{\omega}}$. This is distinct from the sample size of the experiment J, however these simulations set $I = J = 79$ to include all units in the Bloom et al. (2015) dataset.

[13] Our code rounds PJ to the nearest integer value. Note that for the τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR} to be precisely calibrated, the effective $\tilde{P}$ (where $\tilde{P} = \text{round}(PJ)/J$) needs to equal the actual parameter value P.

[14] As with the AR(1) simulations, we set $m = r$ only for simplicity. However, the results are very similar if we fix $m = 3$ and vary $r \in \{1, \ldots, 20\}$, or vice versa.

simulations where $\hat{\tau}^{McK}$, $\hat{\tau}^{AR(1)}$, and $\hat{\tau}^{SCR}$ reject the null hypothesis of $\tau = 0$ at significance level α, under CRVE standard errors. Figure 3.6 reports these three rejection rates on the vertical axes, with the number of pre- and post-treatment periods ($m = r$) on the horizontal axes. We can interpret these curves as realized statistical power, just as in the top row of Figure 3.1. We fix $\alpha = 0.05$ and $\kappa = 0.80$ across all simulations, for the reasons discussed above. Besides our arbitrary choices of $P = 0.5$, all other parameters are determined by the Bloom et al. (2015) dataset: $J = I = 79$, $\sigma_\upsilon^2 = 0.243$, $\sigma_\delta^2 = 0.146$, and $\sigma_\omega^2 = 0.507$, implying $\rho_\upsilon = 0.271$ and $\rho_\delta = 0.163$. We do not estimate β, μ_υ, or μ_δ, as these parameters are no longer relevant when simulating on top of an existing dataset.

B.2.3 Pecan Street data

In Figure 3.8, we present analogous Monte Carlo results for simulations using the Pecan Street dataset of household electricity consumption (Pecan Street 2016). These data are publicly available (with a researcher login) at `https://dataport.pecanstreet.org/data/interactive`, and they include 699 households over 26,888 hours. As with the Bloom et al. (2015) simulations, we construct a balanced panel of households and hours, by restricting the full Pecan Street dataset to a sample of households that report non-missing, non-zero electricity consumption for every hour between January 1, 2013 and December 31, 2014. This results in a balanced panel of $J = 97$ households over 17,520 hours, which we collapse to create daily, weekly, and monthly datasets. Table B.2 presents basic summary statistics for all four datasets, and Figure B.1 displays the time series of data for one randomly selected household in our sample at varying levels of aggregation. Unsurprisingly, while mean electricity use is consistent across different collapses of the data, the standard deviation decreases as we move from higher- to lower-frequency datasets.

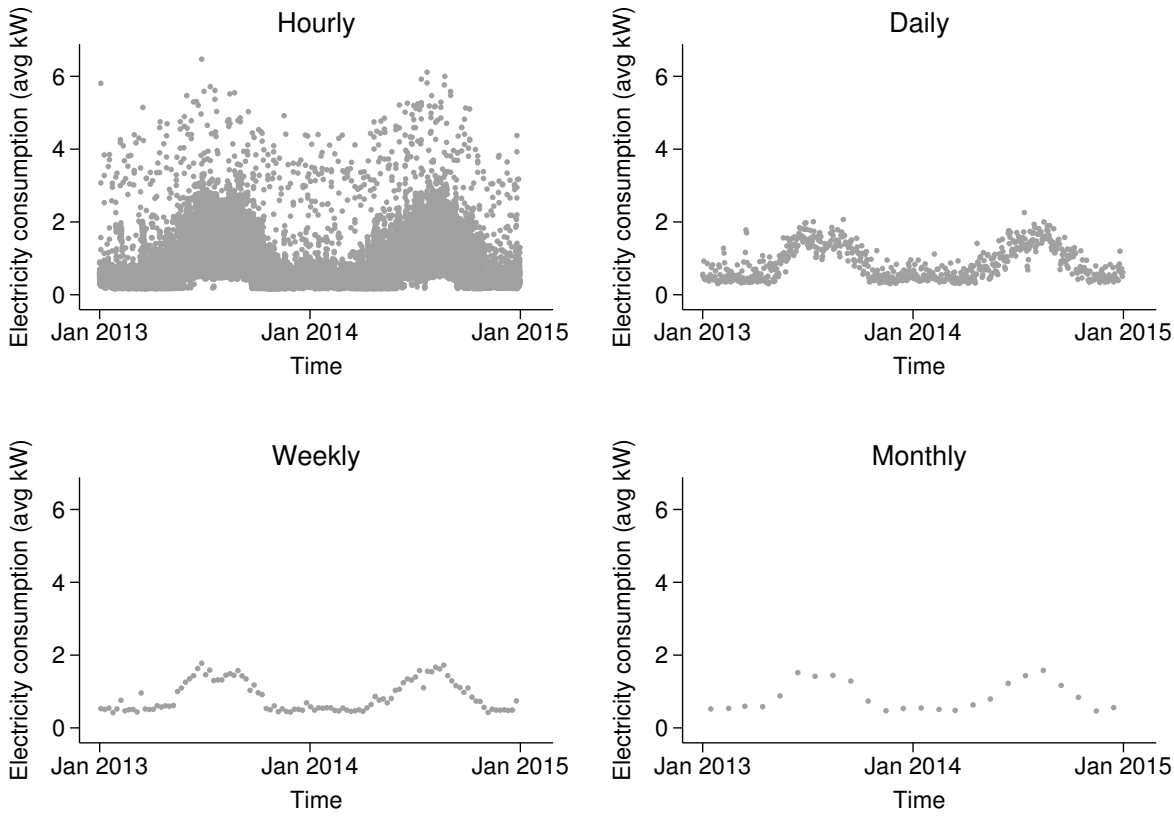
Table B.2: Summary statistics – Pecan Street

Dataset	Mean	Std. Dev.	Min	Max	AR(1) $\hat{\gamma}$	Households	Periods	Observations
Hourly	1.200	1.164	0.019	13.501	0.628	97	17,520	1,699,440
Daily	1.200	0.789	0.082	6.013	0.651	97	730	70,810
Weekly	1.198	0.739	0.105	5.175	0.713	97	106	10,282
Monthly	1.197	0.712	0.169	4.296	0.654	97	24	2,328

Notes: This table shows summary statistics for electricity consumption in the Pecan Street data. All values are in average kW of electricity consumed. The raw data are at the hourly level, in kWh. To construct the daily, weekly, and monthly dataset, we average hourly kWh consumption data across the relevant time period. We compute $\hat{\gamma}$ by estimating Equation (3.6) on residuals from this dataset. In doing so, we cluster standard errors at the household level. All are statistically significant at less than the 1% level.

These Pecan Street simulations follow an algorithm identical to the Bloom et al. (2015) simulations, and we describe this algorithm in detail above. We repeat the same set of simulations four times, estimating separate rejection rates for τ^{McK}, $\tau^{AR(1)}$, and τ^{SCR}, for

Figure B.1: Pecan Street data – Varying levels of aggregation



Notes: This figure shows the time series of Pecan Street electricity consumption data for one randomly selected household in our sample. Each panel displays the data at a different level of aggregation. The data are in units of average kW. These data are highly serially correlated: when we estimate an AR(1) model, Equation (3.6), on residuals from these datasets, we recover the AR(1) parameters of 0.628, 0.651, 0.713, and 0.654 for the hourly, daily, weekly, and monthly data, respectively.

each of the hourly, daily, weekly, and monthly datasets. We again set $\alpha = 0.05$, $\kappa = 0.80$, and $P = 0.5$. The other relevant parameters for each dataset are:

In Appendix B.3.1, we present simulations analogous to Figures 3.2 and 3.3 in the main text, but using the Bloom et al. (2015) and Pecan Street datasets. The former vary m and r separately; the latter replace fixed effects with dummies in the DD estimating equations in Step 4 of Appendix B.2.2.

B.2.4 Analytic power calculations

Figure 3.9 displays the results of analytic power calculations performed using the daily Pecan Street dataset. In other words, we calculate the number of units needed by applying the McKenzie and SCR power calculation formulas, using the data to estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and

Table B.3: Pecan Street Simulation Parameters

Dataset	J	σ^2_υ	σ^2_δ	σ^2_ω	γ	ρ_υ	ρ_δ
Hourly	97	0.257	0.458	0.642	0.623	0.189	0.337
Daily	97	0.257	0.234	0.135	0.651	0.411	0.373
Weekly	97	0.256	0.211	0.083	0.713	0.465	0.384
Monthly	97	0.256	0.203	0.058	0.654	0.495	0.392

These values are estimated separately from each dataset used in the simulations.

$\psi^X_{\hat{\omega}}$. For each experiment of length $m = r \in \{1, \ldots, 12\}$, we estimate the average values of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ over all possible panels of that length.[15] We assign half of the households to treatment ($P = 0.5$), allow for a 5 percent Type I error rate ($\alpha = 0.05$), and calibrate to 80 percent power ($\kappa = 0.80$). Finally, we rearrange Equations (3.3) and (B.86) to calculate the number of households required to detect MDEs that range from 0 to 15 percent of baseline electricity consumption.

Figure 3.10 also shows the results of analytic power calculations using the SCR formula. However, instead of parameterizing Equation (3.2) using estimates from a dataset, we now normalize $\sigma^2_\omega = 1$ and assume an AR(1) correlation structure with $\gamma \in \{0, 0.3, 0.5, 0.7, 0.9\}$. For panel lengths $m = r \in \{1, \ldots, 100\}$, we analytically derive ψ^B, ψ^A, and ψ^X using the formulas from Equations (B.81)–(B.83). In the left panel, we fix $P = 0.5$, $\alpha = 0.05$, $\kappa = 0.80$, and $J = 100$, and use Equation (3.2) to solve for MDE as a function of $m = r$ and γ. In the right panel, we fix $P = 0.5$, $\alpha = 0.05$, $\kappa = 0.80$, and $MDE = 1$, and rearrange Equation (3.2) to solve for J as a function of $m = r$ and γ.

B.3 Additional results

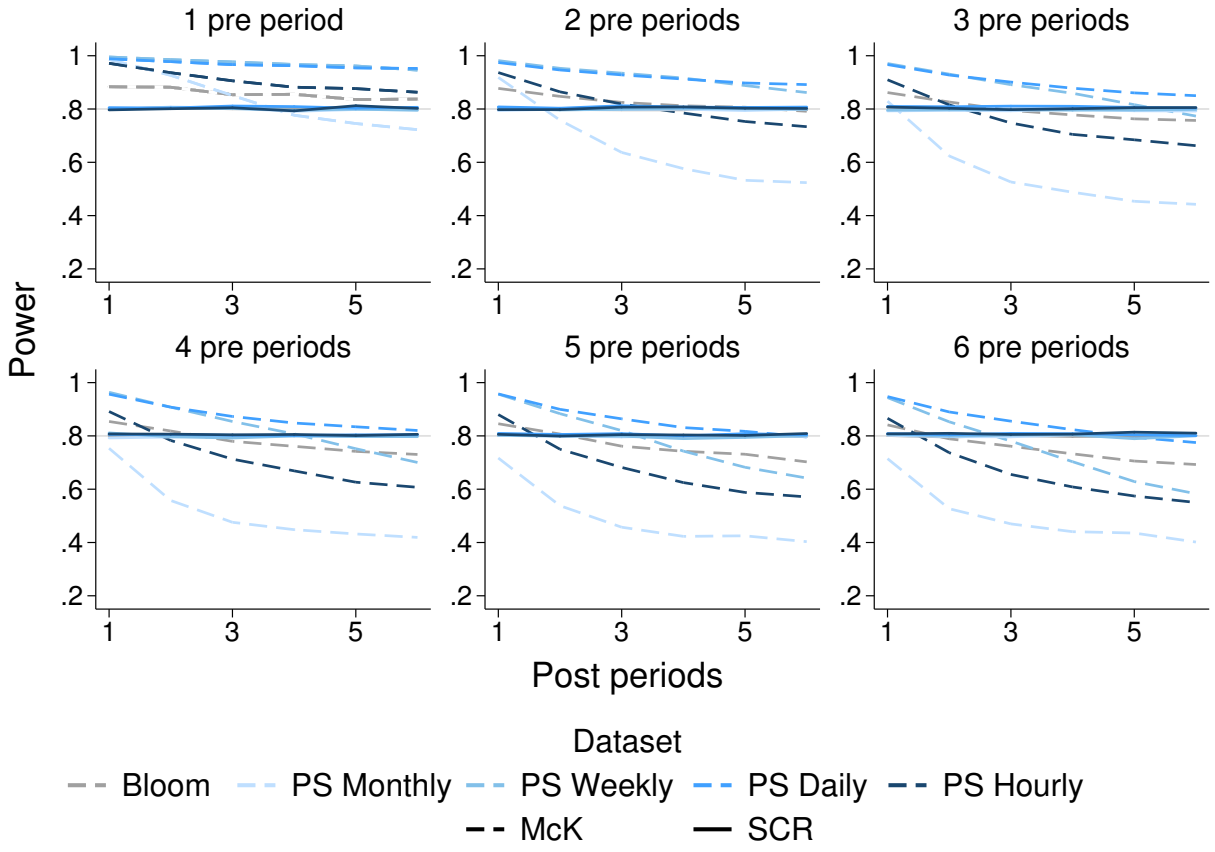
B.3.1 Sensitivities with real data

In this section, we present extensions of our simulation results from the main text. In Figure B.2, we present an analogous version of Figure 3.2 using real data from Bloom et al. (2015) and Pecan Street. Each panel conducts simulations that are identical to Figures 3.6 and 3.8, varying m and r separately (for $m \in \{1, \ldots, 6\}$ and $r \in \{1, \ldots, 6\}$). As with the simulated data, we find that the McKenzie formula typically yields over-powered experiments with either one pre-treatment or one post-treatment period, while the SCR formula yields the

[15] We follow the algorithm outlined in Appendix B.4.1 below. For a given value of $m = r$, we consider each (consecutive) subset S of the daily Pecan Street data with length $2r$. We first residualize this subset of the data with household and day fixed effects, and we calculate $\sigma^2_{\hat{\omega},S}$ from these residuals. We then assign the first m residuals for each household to the pre-treatment period and the remaining r residuals to the post-treatment period, thereby estimating $\psi^B_{\hat{\omega},S}$, $\psi^A_{\hat{\omega},S}$, and $\psi^X_{\hat{\omega},S}$ (by averaging all pairwise covariances for subset S). Averaging $\sigma^2_{\hat{\omega},S}$, $\psi^B_{\hat{\omega},S}$, $\psi^A_{\hat{\omega},S}$, and $\psi^X_{\hat{\omega},S}$ over all subsets S, we arrive at estimates for $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$.

desired 80 percent power in all cases.

Figure B.2: Power in short panels – Real data

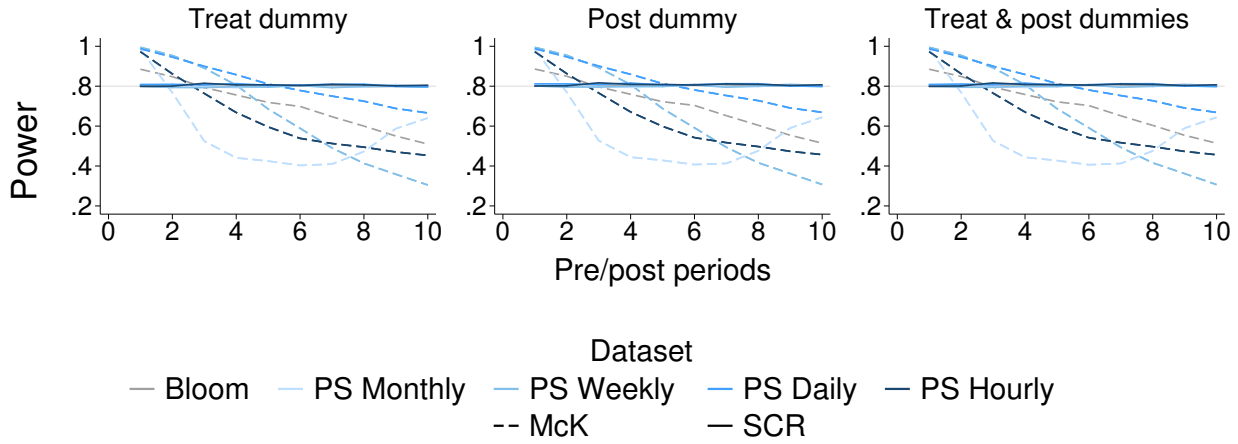


Notes: This figure combines Figures 3.6 and 3.8 from the main text with the short panels of Figure 3.2. Simulations using the Bloom et al. (2015) and Pecan Street datasets follow the same algorithm describe in Appendix B.2.2, but separately varying the number of pre-treatment and post-treatment periods (for $m \in \{1, \ldots, 6\}$ and $r \in \{1, \ldots, 6\}$). For nearly all cases with either one pre-treatment period or one post-treatment period, the McKenzie formula yields over-powered experiments. Hence, experiments that follow the traditional "one baseline, one follow-up" structure will likely be overpowered, having calibrated an excessively large sample size. As the number of pre-treatment periods increases, power decreases monotonically for each dataset. By contrast, the SCR formula is properly powered in all cases.

Next, Figure B.3 presents an analogous version of Figure 3.3 from the main text, using Bloom et al. (2015) and Pecan Street datasets. In the left panel, we replace unit fixed effects in the DD estimating equation with a Treat_i dummy. In the middle panel, we replace time fixed effects in the DD estimating equation with a Post_t dummy. In the right panel, we replace both unit and time fixed effects with $\text{Treat}_i + \text{Post}_t$ dummies. These resulting rejection rates are virtually identical to Figures 3.6 and 3.8, demonstrating that these alternative

DD estimating equations all yield identical *ex post* power in real data (just as they do with simulated data in Figure 3.3).

Figure B.3: Sensitivities to estimating equation – Real data



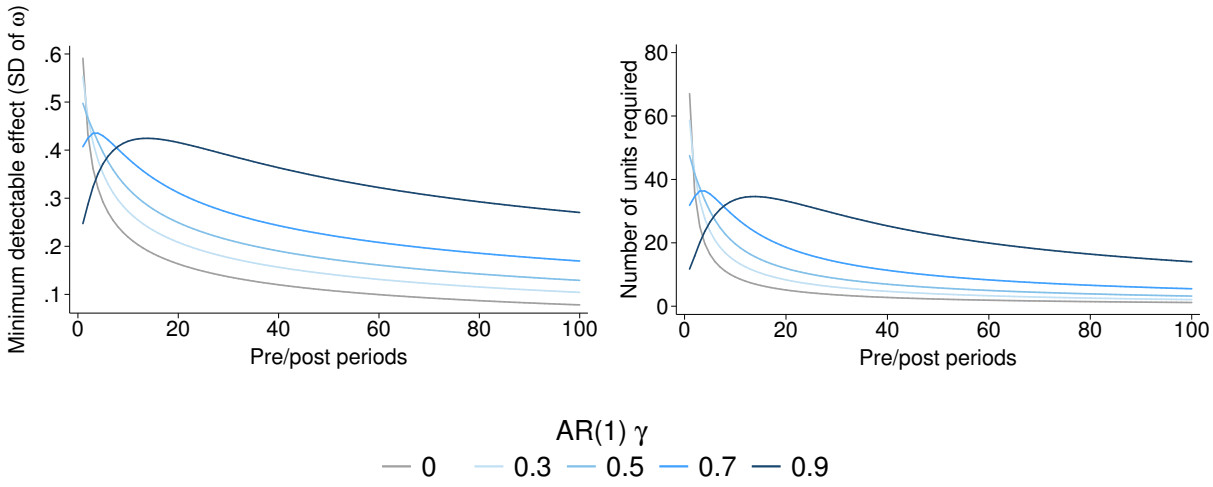
Notes: This figure replicates Figure 3.3 from the main text, using the Bloom et al. (2015) and Pecan Street datasets. The left panel replaces unit fixed effects in the DD estimating equation with a Treat_i dummy. The middle panel replaces time fixed effects in the DD estimating equation with a Post_t dummy. The right panel replaces both unit and time fixed effects with $\text{Treat}_i + \text{Post}_t$ dummies. As with simulated data, DD power calculations using real data are incorrectly powered using the McKenzie formula but correctly powered using our SCR formula.

B.3.2 Tradeoffs between MDE and time periods: ANCOVA

Figure 3.10 demonstrates that for DD experiments with strong serial correlation, it is possible to increase the MDE (or necessary sample size) by adding pre/post-treatment periods. This follows from Equation (3.4), which shows that serial correlation can either increase or decrease the variance of the DD estimator. Because DD identifies the treatment effect using differences between pre- and post-treatment outcomes, stronger serial correlation makes differences caused by treatment easier to detect. At the same time, stronger serial correlation means each additional time period provides less information. This intuition also holds for the ANCOVA estimator, as shown by Figure B.4. This replicates Figure 3.10 using the SCR ANCOVA power calculation formula (Equation (3.11)). As with DD, we find that adding pre/post periods to an ANCOVA experiment can decrease power, for short panels with strong serial correlation.[16]

[16]The SCR ANCOVA formula has an additional degree of freedom: we can vary σ_v^2 and σ_ω^2 separately. Here, we set a relatively small σ_v^2 (equal to $0.1\sigma_\omega^2$), which causes ANCOVA to be relatively dissimilar to DD (the two estimators converge as σ_v^2 approaches infinity).

Figure B.4: Analytical power calculations with increasing panel length – ANCOVA



Notes: This figure replicates Figure 3.10 from the main text, using the SCR ANCOVA formula (Equation (3.11)) instead of the SCR DD formula (Equation (3.2)). The left panel shows the tradeoff between the minimum detectable effect (MDE) and the number of time periods ($m = r$) for varying levels of AR(1) serial correlation, holding the number of units fixed at $J = 100$ and normalizing MDE by the standard deviation of ω_{it}. At low levels of γ, MDE declines monotonically in m and r. However, for higher γ, increasing m and r actually *increases* MDE when $m = r$ is relatively small, and *decreases* MDE when $m = r$ is relatively large. The right panel shows the relationship between the number of units (J) and number of pre/post periods ($m = r$) required to detect an MDE equal to one standard deviation of ω_{it}. Similarly, for low levels of serial correlation, the trade-off between J and $m = r$ is monotonic. However, as γ increases, adding periods in short panels necessitates a greater number of units to achieve the same MDE, while adding periods in longer panels means that fewer units are required to achieve the same MDE.

B.3.3 Cluster randomization in panel RCTs

Finally, we use simulation-based power calculations to compare panel RCTs with unit-level randomization and cluster-level randomization. While cluster randomization remains outside the scope of this paper's analytical framework, it is a common approach for RCTs in development economics. Randomizing at the cluster (e.g. village) level, rather than the unit (e.g. household) level can be less expensive, simplify the logistics of administering treatment, and eliminate concerns about treatment interference (e.g. spillovers from treated to control households within a village). Here, we investigate how cluster-level randomization affects statistical power (relative to unit-level randomization), using the program `pc_simulate` from our STATA package `pcpanel`.[17]

We compare unit-level vs. cluster-level randomization, using simulated panel datasets based on the following DGP:

[17]We do not incorporate treatment spillovers, since doing so in a simulation context would necessitate (essentially) arbitrary assumptions on the strength and direction of treatment spillovers. Instead, we simulate RCTs with a 100 percent treatment intensity within treated clusters.

Assumption B22 (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{igt} = \beta + \tau D_{igt} + \zeta_g + \upsilon_i + \delta_t + \omega_{igt} \tag{B.84}$$

where Y_{igt} is the outcome of interest for unit i in group g at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{igt} is a time-varying treatment indicator; ζ_g is a group-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\zeta)$; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\upsilon)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma^2_\delta)$; and ω_{igt} is an idiosyncratic error term with variance σ^2_ω.

We simulate 9 datasets, each containing 500 groups g, where each group is defined by a separate realization of ζ_g. We vary the number of units i nested within each group g, such that all groups within a given dataset comprise either 10, 25, or 50 units, but hold the total number of units fixed. We also vary the relative strength of group-specific shocks while holding constant total cross-sectional variation—that is, we vary $\sigma^2_\zeta/(\sigma^2_\zeta + \sigma^2_\upsilon)$ such that groups contribute 25, 50, of 75 percent of the total cross-sectional variation, while holding $\sigma^2_\zeta + \sigma^2_\upsilon$ fixed. For all datasets, we incorporate non-constant serial correlation where ω_{igt} follows an AR(1) process with $\gamma = 0.5$.[18]

Using `pc_simulate`, we conduct 18 sets of simulation-based power calculations, for varying panel lengths ($m = r \in \{1, 2, 3, 4, 5, 6\}$). While all 18 sets of power calculations hold the sample size fixed at 2000 units i, 9 assign treatment at the unit level (ignoring cross-sectional group correlations) and 9 assign treatment at the group level (with 100 percent treatment intensity within treated groups). Hence, for group sizes of 10, 25, and 50, this means that cluster randomization selects sample sizes of 200, 80, and 40 groups, respectively. We also hold the proportion of units in treatment and minimum detectable effect constant across all simulations ($P = 0.5$, $MDE = 1.4$). All simulations cluster standard errors at the level of treatment (either unit or group), and calculate average rejection rates over 10,000 iterations to compute power.[19]

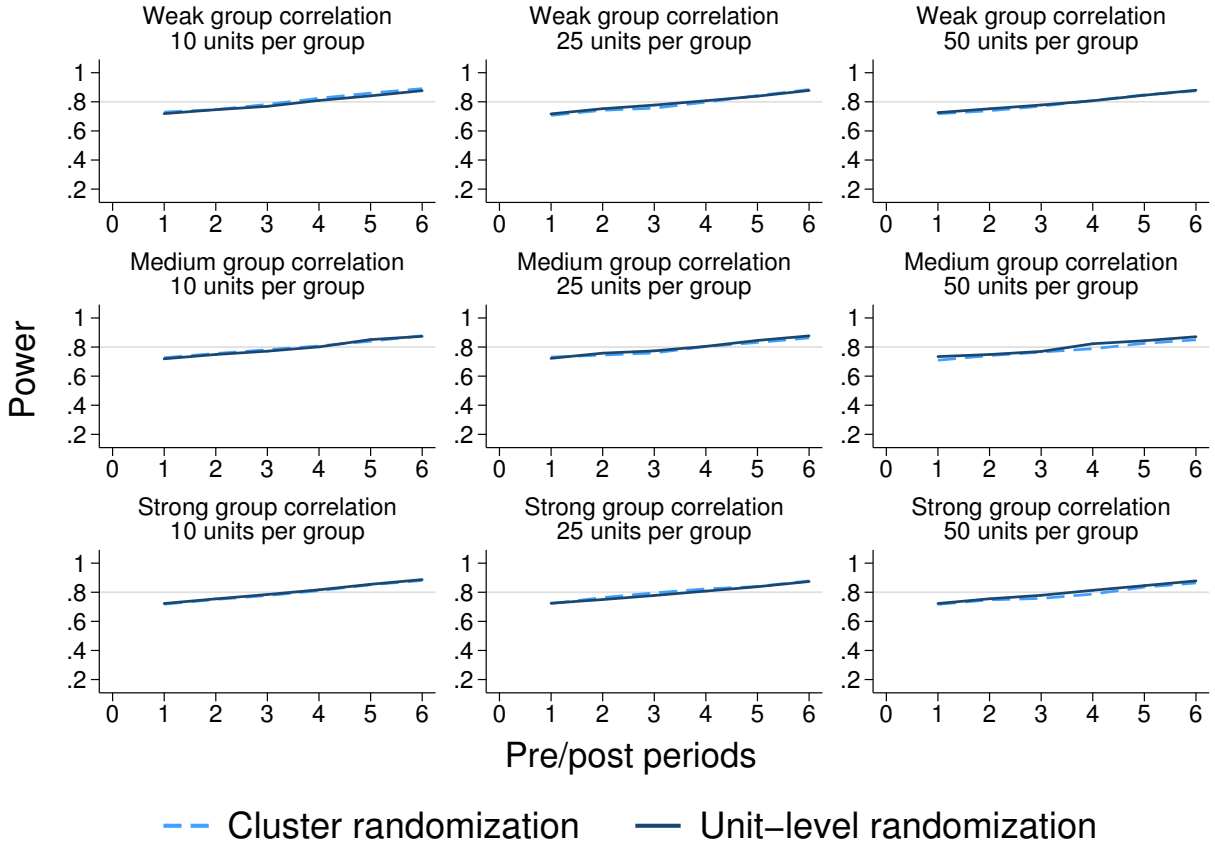
Figure B.5 presents the results of this exercise. Each plot reports an "apples-to-apples" comparison for power calculations on the same dataset, randomized at the unit vs. group level. Moving left to right, the number of units per group increases from 10 to 25 to 50. Moving top to bottom, the relative strength of group-specific shocks increases from 25 percent ("weak") to 50 percent ("medium") to 75 percent ("strong"). For all 9 datasets, randomizing at the group level yields statistical power that is indistinguishable from the same experiment randomized at the unit level. The intuition behind this equivalence lies in how panel estimators control for unit-specific baselines: DD identifies off of (post − pre) differences at the *unit* level, which controls for the remaining within-group variation not exploited when

[18] We set the following specific parameters: $\sigma^2_\zeta \in \{20, 40, 60\}$, $\sigma^2_\upsilon \in \{60, 40, 20\}$, $\sigma^2_\zeta + \sigma^2_\upsilon = 80$, $\sigma^2_\delta = 10$, $\sigma^2_\omega = 150$. All parameter values are arbitrary, besides imposing constant $\sigma^2_\zeta + \sigma^2_\upsilon$ (which facilitates comparisons across datasets).

[19] We use a large sample size of 2000 units to ensure that for datasets with 50-unit groups, we still have a sufficient number of clusters (40) to apply the CRVE.

randomizing at the group level. By contrast, when we repeat the same exercise for cross-sectional RCTs, we finds substantially lower power for group-level randomization (relative to unit-level randomization). This is because the cross-sectional treatment effect estimator does not control for within-group variation, meaning that the treatment indicator leaves more unexplained variation when randomized at the group level. Because these results are on simulations in fake data with a relatively simple DGP, we encourage researchers interested in performing cluster randomization to do so using simulation-based power calculations, in case these results do not fully translate to real-world datasets with more complex DGPs.

Figure B.5: Power calculations by simulation – Cluster randomization



Notes: This figure reports results from simulation-based power calculations, comparing unit-level randomization vs. cluster randomization. Each panel reports power for DD experiments using a single simulated dataset, varying the number of pre-treatment and post-treatment periods ($m = r = \{1, 2, 3, 4, 5, 6\}$). Datasets have either 10, 25, or 50 units within each group (moving left to right); and either weak, medium, or strong within-group correlations (either 25, 50, or 75 percent of total cross-sectional variation, respectively). All experiments include 2000 units, treating either half of units or half of groups. When cluster randomizing, all treated groups have a treatment intensity of 100 percent. All simulations assume an MDE of 1.4, and calculate average rejection rates over 10,000 iterations. See text for further details.

B.4 A practical guide to power calculations

In this section, we address several practical considerations when conducting power calculations. Most of these challenges involve variance and covariance parameters that must be either estimated or assumed in order to operationalize a power calculation formula. We also outline steps for estimating power calculations via simulation, which is our preferred method.

B.4.1 Analytical power calculations

The most challenging aspect of analytical power calculations is parameterizing the variance and (if applicable) covariance terms that characterize the data's error structure. In the absence of a representative pre-existing dataset, researchers may struggle to even guess the order of magnitude of the error variance, let alone generate a precise estimate of this key parameter. Our theoretical results demonstrate that power for panel RCTs also hinges on the error *covariance* structure.

As in the paper, we denote the true parameters governing the data generating process as σ^2_ω, ψ^B, ψ^A, and ψ^X. We define $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ to be the parameters that characterize the *residuals* (rather than real errors). If researchers have access to a representative dataset *ex ante*, they can directly estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, and use these values to parameterize power calculations. Appendix B.5.2 proves that researchers can recover the true MDE using these residual-based parameters. Nevertheless, this process is not trivial, for several reasons.

First, while the idiosyncratic variance σ^2_ω is a population parameter, the three ψ parameters are functions of both the full covariance structure of the population and the specific values of m and r.[20] For a given population and serially correlated outcome variable, experiments with small m and r are likely to exhibit larger ψ^B, ψ^A, and ψ^X parameters than experiments with large m and r. This is because as the number of pre-treatment (post-treatment) periods increases, ψ^B (ψ^A) averages across covariances of time periods that are farther apart. For example, compare ψ^B with $m = 3$ vs. $m = 30$, for an outcome with a covariance structure where adjacent periods are more positively correlated than distant periods. For $m = 3$, ψ^B averages $m(m-1)/2 = 3$ pairwise covariances, 2 of which are for adjacent periods; for $m = 30$, ψ^B averages $m(m-1)/2 = 435$ pairwise covariances, only 29 of which are for adjacent periods. Because ψ^X expands with both m and r, it attenuates relatively faster than ψ^B and ψ^A as panel length grows.

Second, estimating $\text{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is})$ using residuals from an existing dataset is fundamentally impossible, given that each dataset contains only one realization of (Y_{it}, Y_{is}). However, researchers may treat the $(I \times 1)$ vectors of residuals $(\hat{\boldsymbol{\omega}}_t, \hat{\boldsymbol{\omega}}_s)$ as I draws from the distributions of residuals for periods (t,s) and estimate these distributions' covariance. The resulting estimates, which we denote $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$, are unbiased estimators of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$.[21]

Third, if the representative dataset contains a long time series, the residual variance and covariance structure may change throughout the time series. This means if researchers estimate $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ by averaging across the full time series, these estimated parameters may be less representative than if they were estimated from just the end of the time series.[22] Since the residual variance is not a function of panel length, it may be tempting to estimate $\tilde{\sigma}^2_{\hat{\omega}}$ using a long vector of residuals, while estimating $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ using only residuals

[20] Deriving the residual-based parameters $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ introduces an additional complexity, as these residual-based parameters are defined by the number of pre-treatment periods (m), post-treatment periods (r) and cross-sectional units (I) used to produce these residuals.

[21] Appendix B.5.1 proves that $\mathbb{E}[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}] = \sigma^2_{\hat{\omega}}$, $\mathbb{E}[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}] = \psi^B_{\hat{\omega}}$, $\mathbb{E}[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}] = \psi^A_{\hat{\omega}}$, and $\mathbb{E}[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}] = \psi^X_{\hat{\omega}}$.

[22] Of course, if the researcher expects a certain subset of her data is likely more representative, it would be wise to perform power calculations on this subset alone.

within an $(m+r)$-period range. However, in a time series where the variance-covariance structure is changing, this would produce $\tilde{\psi}_{\omega}$ estimates that are inconsistent with $\tilde{\sigma}^2_{\hat{\omega}}$.

Fourth, while $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ are unbiased estimators of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, they are **not** unbiased estimators of σ^2_{ω}, ψ^B, ψ^A, and ψ^X. This is because the residuals from the regression $Y_{it} = \upsilon_i + \delta_t + \omega_{it}$ will have a variance less than the parameter σ^2_{ω} from the data generating process, by the properties of linear projection. In addition, when they are estimated using residuals from shorter panels, $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ have a more severe bias, but these estimates converge to their true values (i.e., σ^2_{ω}, ψ^B, ψ^A, and ψ^X) as the panel length used to estimate these residuals increases.[23] Importantly, for the purposes of power calculations using the SCR formula, we *can* recover an unbiased estimate of the minimum detectable effect with the true parameters using our parameter estimates. That is, $MDE^{est}(\sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}) = MDE(\sigma^2_{\omega}, \psi^B, \psi^A, \psi^X)$. As Appendix B.5.1 shows, $\mathbb{E}[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}] = \sigma^2_{\hat{\omega}}$, $\mathbb{E}[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}] = \psi^B_{\hat{\omega}}$, $\mathbb{E}[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}] = \psi^A_{\hat{\omega}}$, and $\mathbb{E}[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}] = \psi^X_{\hat{\omega}}$. Combining these two proofs suggests that $MDE^{est}(\mathbb{E}[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}], \mathbb{E}[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}], \mathbb{E}[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}], \mathbb{E}[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}]) = MDE(\sigma^2_{\omega}, \psi^B, \psi^A, \psi^X)$. Therefore, for values of $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ estimated from a pre-existing dataset with I cross-sectional units:

$$
\begin{aligned}
MDE^{est} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \Bigg\{ & \left(\frac{1}{P(1-P)J}\right) \\
& \times \left[\left(\frac{m+r}{mr}\right) k_{\sigma}\, \mathbb{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right] + \left(\frac{m-1}{m}\right) k_B\, \mathbb{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right]\right. \\
& \left. + \left(\frac{r-1}{r}\right) k_A\, \mathbb{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right] - 2k_X\, \mathbb{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right]\right] \Bigg\}^{1/2} \qquad \text{(B.85)} \\
= \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \Bigg\{ & \left(\frac{1}{P(1-P)J}\right) \left[\left(\frac{m+r}{mr}\right) k_{\sigma}\sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right) k_B\psi^B_{\hat{\omega}}\right. \\
& \left. + \left(\frac{r-1}{r}\right) k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}}\right] \Bigg\}^{1/2} \qquad \text{(B.86)}
\end{aligned}
$$

where

$$
\begin{aligned}
k_{\sigma} &= \frac{I(m+r)^2}{2(I-1)mr} \\
k_B &= \frac{I(m+r)^2}{2(I-1)r^2} \\
k_A &= \frac{I(m+r)^2}{2(I-1)m^2} \\
k_X &= 0
\end{aligned}
$$

and the expectation of parameters are taken over subsets of the dataset, as described in the next point. Appendix B.5.2 proves that Equations (B.85) and (B.86) are equivalent, and derives the above expressions for the coefficients k_{σ}, k_B, k_A, and k_X.

[23] The estimated residuals include both the true idiosyncratic error, ω_{it}, and (attenuating) fixed-effect estimation error. Although both sets of fixed effects, υ_i and δ_t, are unbiased and consistent in T and I, respectively, error in estimating these parameters will always yield residuals that are smaller on average, biasing the estimation of these parameters. The estimation error and resulting biases decrease in T and I.

Fifth, because the estimated variance-covariance terms enter the power calculation under a radical, researchers must be conscious of Jensen's Inequality. If the researcher is estimating $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ by taking the expectation of $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ across a range of $(m+r)$-period subsets, then the correct calculation is:

$$MDE^{est}\left(\mathbb{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right], \mathbb{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right], \mathbb{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right], \mathbb{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right]\right), \text{ not } \mathbb{E}\left[MDE^{est}\left(\tilde{\sigma}^2_{\hat{\omega}}, \tilde{\psi}^B_{\hat{\omega}}, \tilde{\psi}^A_{\hat{\omega}}, \tilde{\psi}^X_{\hat{\omega}}\right) \mid \mathbf{X}\right].$$

Similarly, if Equation (3.2) is rearranged as a function of κ, it becomes convex in the variance-covariance parameters, and the correct calculation is:

$$\kappa\left(\mathbb{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right], \mathbb{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right], \mathbb{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right], \mathbb{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right]\right), \text{ not } \mathbb{E}\left[\kappa\left(\tilde{\sigma}^2_{\hat{\omega}}, \tilde{\psi}^B_{\hat{\omega}}, \tilde{\psi}^A_{\hat{\omega}}, \tilde{\psi}^X_{\hat{\omega}}\right) \mid \mathbf{X}\right].$$

When solving for sample size J, Equation (3.2) becomes linear in variance-covariance parameters, meaning that Jensen's Inequality does not affect the estimate of $J\left(\sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}\right)$.

In light of each of these issues, we recommend the following algorithm for estimating the MDE using a pre-existing panel dataset (implemented by our STATA program `pc_dd_covar`):

1. Determine all feasible ranges of experiments with $(m+r)$ periods, given the number of time periods in the pre-existing dataset. For example, if this dataset contains 100 time periods indexed $t = \{1, \ldots, 100\}$, and $m = 5$ and $r = 6$, then there are 90 feasible ranges for an experiment with $(m+r) = 11$ periods (i.e., beginning in periods $t = \{1, \ldots, 90\}$).

2. For each feasible range S:

   a) Regress the outcome variable on unit and time-period fixed effects, $Y_{it} = \upsilon_i + \delta_t + \omega_{it}$, and store the residuals. (This regression includes all I available cross-sectional units, but only time periods with the specific range S.[24])

   b) Calculate the variance of the stored residuals, and save as $\tilde{\sigma}^2_{\hat{\omega},S}$.

   c) For each pair of pre-treatment periods (i.e., the first m periods in range S), calculate the covariance between these periods' residuals. Take an unweighted average of these $m(m-1)/2$ covariances, and save as $\tilde{\psi}^B_{\hat{\omega},S}$.

      For example, if $m = 4$, $r = 2$, and range S begins in period $t = 1$, sum $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$, $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_3)$, $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_4)$, $\text{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_3)$, $\text{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_4)$, and $\text{Cov}(\boldsymbol{\omega}_3, \boldsymbol{\omega}_4)$, and divide by $m(m-1)/2 = 6$.

   d) For each pair of post-treatment periods (i.e., the last r periods in range S), calculate the covariance between these periods' residuals. Take an unweighted average of these $r(r-1)/2$ covariances, and save as $\tilde{\psi}^A_{\hat{\omega},S}$.

      For example, if $m = 4$, $r = 2$, and range S begins in period $t = 1$, $\tilde{\psi}^A_{\hat{\omega},S}$ is the average of a single post-period covariance, $\text{Cov}(\boldsymbol{\omega}_5, \boldsymbol{\omega}_6)$.

[24] This bears no relationship to the sample size J units to be included in the power calculation. Assuming that all I units in the pre-existing dataset represent the population to be included in the randomization, estimating the variance and covariances using all available units will provide the best estimates of $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$ (by the Weak Law of Large Numbers).

e) For each pair of pre- and post-treatment periods (i.e. the first m and the last r periods in range S), calculate the covariance between these periods' residuals. Take an unweighted average of these mr covariances, and save as $\tilde{\psi}^X_{\hat{\omega},S}$.

For example, if $m = 4$, $r = 2$, and range S begins in period $t = 1$, sum $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_5)$, $\text{Cov}(\boldsymbol{\omega}_1, \boldsymbol{\omega}_6)$, $\text{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_5)$, $\text{Cov}(\boldsymbol{\omega}_2, \boldsymbol{\omega}_6)$, $\text{Cov}(\boldsymbol{\omega}_3, \boldsymbol{\omega}_5)$, $\text{Cov}(\boldsymbol{\omega}_3, \boldsymbol{\omega}_6)$, $\text{Cov}(\boldsymbol{\omega}_4, \boldsymbol{\omega}_5)$, and $\text{Cov}(\boldsymbol{\omega}_4, \boldsymbol{\omega}_6)$, and divide by $mr = 8$.

3. Calculate the average of $\tilde{\sigma}^2_{\hat{\omega},S}$, $\tilde{\psi}^B_{\hat{\omega},S}$, $\tilde{\psi}^A_{\hat{\omega},S}$, and $\tilde{\psi}^X_{\hat{\omega},S}$ across all ranges S, deflating $\tilde{\sigma}^2_{\hat{\omega},S}$ by $\frac{IT-1}{IT}$, and $\tilde{\psi}^B_{\hat{\omega},S}$, $\tilde{\psi}^A_{\hat{\omega},S}$, and $\tilde{\psi}^X_{\hat{\omega},S}$ by $\frac{I-1}{I}$. These averages are equal in expectation to $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$.

4. Plug these values into Equation (B.86) to produce MDE^{est}.

Figure B.6 highlgihts the difference between true and estimated variance-covariance parameters in AR(1) data. In particular, we show true parameter values of σ^2_{ω}, ψ^B, ψ^A, and ψ^X alongside estimated values of these same parameters, calculated according to the procedure outlined above. As expected, $\sigma^2_{\hat{\omega}}$ is biased downwards relative to σ^2_{ω}, but converges towards this value as the panel length increases. This convergence is slower for larger AR(1) parameters, as highly serially correlated errors make it harder to identify the unit fixed effects. Similarly, while the true ψ^X is positive across all panel lengths, $\psi^X_{\hat{\omega}}$ is negative everywhere, and ψ^B and ψ^A also differ from their estimated counterparts. Despite the differences between the true parameters and their estimated values, Appendix B.5.2 proves that we can recover the MDE based on true underlying parameters using residual-based parameters. In conjunction with the fact that we can estimate the residual-based parameters from real data, this confirms that researchers can use estimated parameters to calibrate power calculations.

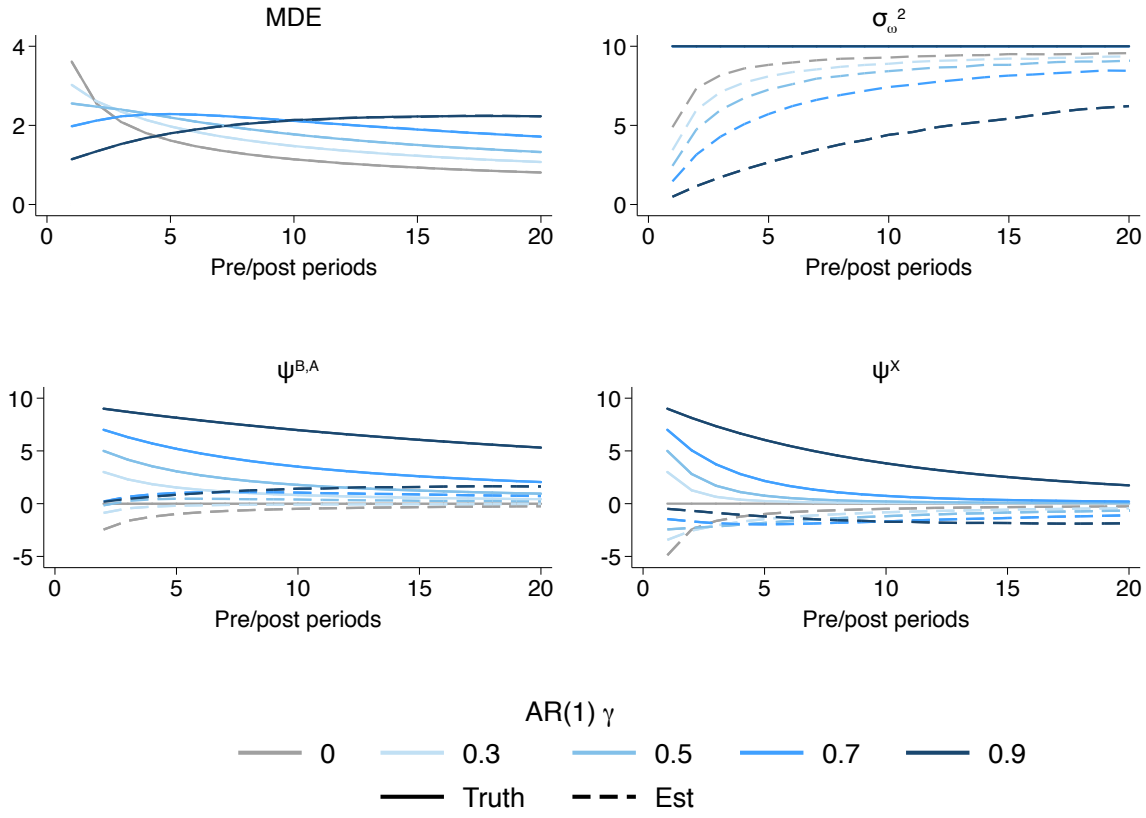
Figure B.7 uses the Bloom et al. (2015) dataset to present an analogous comparison between actual vs. estimated σ^2_{ω}, ψ^B, ψ^A, and ψ^X parameters. Here, as in Figure B.6, the dotted lines estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ using the above algorithm. However, unlike with simulated AR(1) datasets, the "true" parameters of the Bloom et al. (2015) data generating process are unknown. We estimate these "true" values using residuals from the full 48-period time series, which minimizes the fixed effect estimation error that biases $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ in short panels.[25] This reveals a very similar pattern: "subsetted" $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ estimates are systematically biased downward, but converge to their "full time-series" (i.e. closer to "true") values as panel length increases. As in Figure B.6, we show that both sets of estimated variance-covariance parameters yield (virtually) identical MDEs, as long as Equation (3.2) uses estimated parameters that are internally consistent.

Figure B.8 replicates Figure B.7 for all four Pecan Street datasets. We see that while the estimated values $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ differ across different levels of aggregation, they follow the same pattern. The subsetted estimates are biased downward, but appear to converge to the full time series estimates (i.e. closer to truth) as panel length increases. In all four cases,

[25] These $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ estimates (represented by solid lines in Figure B.7) result from the same algorithm as detailed above, except omitting Step 2(a) and estimating a single 48-period set of residuals in Step 1. This provides the closest possible approximation to the "true" variance-covariance structure of these data, and hence the most apples-to-apples comparison to Figure B.6.

the MDE is (virtually) identical when calculated using either all full time series estimates or all subsetted estimates.

Figure B.6: Actual vs. estimated parameters – AR(1) data



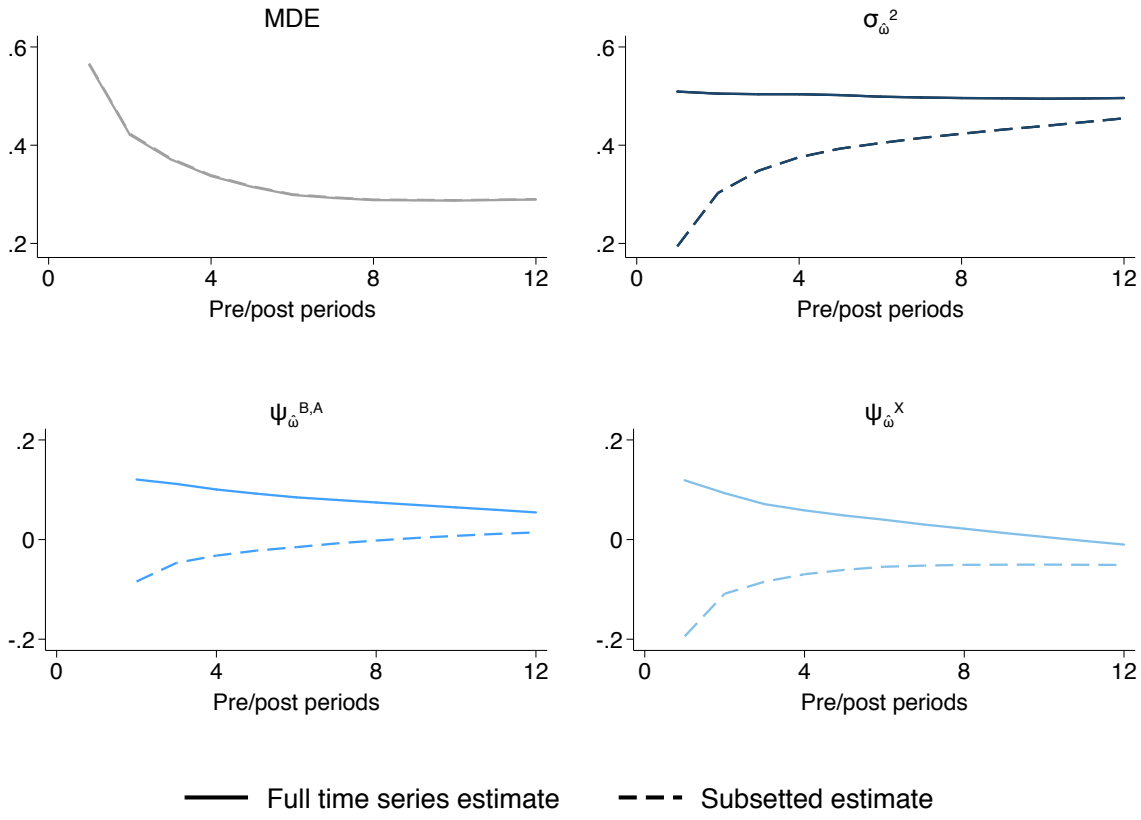
Notes: This figure displays the difference between the true residual variance (σ^2_ω), average pre (post)-period covariance ($\psi^{B,A}$), and average cross-period covariance (ψ^X), and their estimated counterparts over varying panel lengths. It also shows the resulting minimum detectable effect (MDE), calculated using the SCR formula (Equation (3.2)). These parameters and estimates come from a simulated datasets with AR(1) errors, generated identically to those presented in Figure 3.1. True parameters are displayed with solid lines (with ψ terms derived analytical using Equations (B.81)–(B.83)), and estimates are displayed with dashed lines (estimated according to the algorithm described above). As expected, the true σ^2_ω is constant across panel lengths, while $\psi^{B,A}$ declines with the number of pre (post) periods, and ψ^X declines more quickly than $\psi^{B,A}$ as the panel length increases. The higher the AR(1) parameter, the larger are the ψ terms. The estimated parameters behave quite differently from their true counterparts. In short panels, $\sigma^2_{\hat{\omega}}$ is biased downwards, because the regression model's individual fixed effects are inconsistently estimated, and capture more of the true error variance than they should explain in expectation. This has the effect of reducing the estimated covariances $\psi^{B,A}_{\hat{\omega}}$, which scale with $\sigma^2_{\hat{\omega}}$. At the same time, $\psi^X_{\hat{\omega}}$ is mechanically negative, as the estimated fixed effects yield residuals that are negatively correlated within individuals across pre/post-treatment time periods (and some of the variation that should be captured by $\sigma^2_{\hat{\omega}}$ is instead loaded on to the $\psi_{\hat{\omega}}$ terms). As panel length increases, the $\psi_{\hat{\omega}}$ terms converge to the true ψ values. Estimated parameters result in the same MDEs as real parameters after applying the correction factors in Equation (B.86), as demonstrated by the top left panel.

Two additional nuances that arise during analytical power calculations are worth noting. First, the critical values $t^d_{1-\kappa}$ and $t^d_{\alpha/2}$ should be drawn from an inverse t-distribution with the same degrees of freedom as the *ex post* regression model. This means that if researchers plan to use CRVE standard errors clustered by experimental unit, they should draw these critical values from an inverse t-distribution with J degrees of freedom. To be precise, these critical values should be sensitive to changes in the number of unit/clusters J, although the t degrees of freedom has a very small effect on MDE, relative to other parameters in Equation (3.2).[26]

Finally, in panel RCTs with CRVE standard errors clustered by unit, the proportion of units treated P cannot be too large or too small. Our simulations have demonstrated that the Equation (3.2) performs poorly if $P < 0.1$ or $P > 0.9$, because the CRVE requires a sufficient number of clusters that are both treated and control.

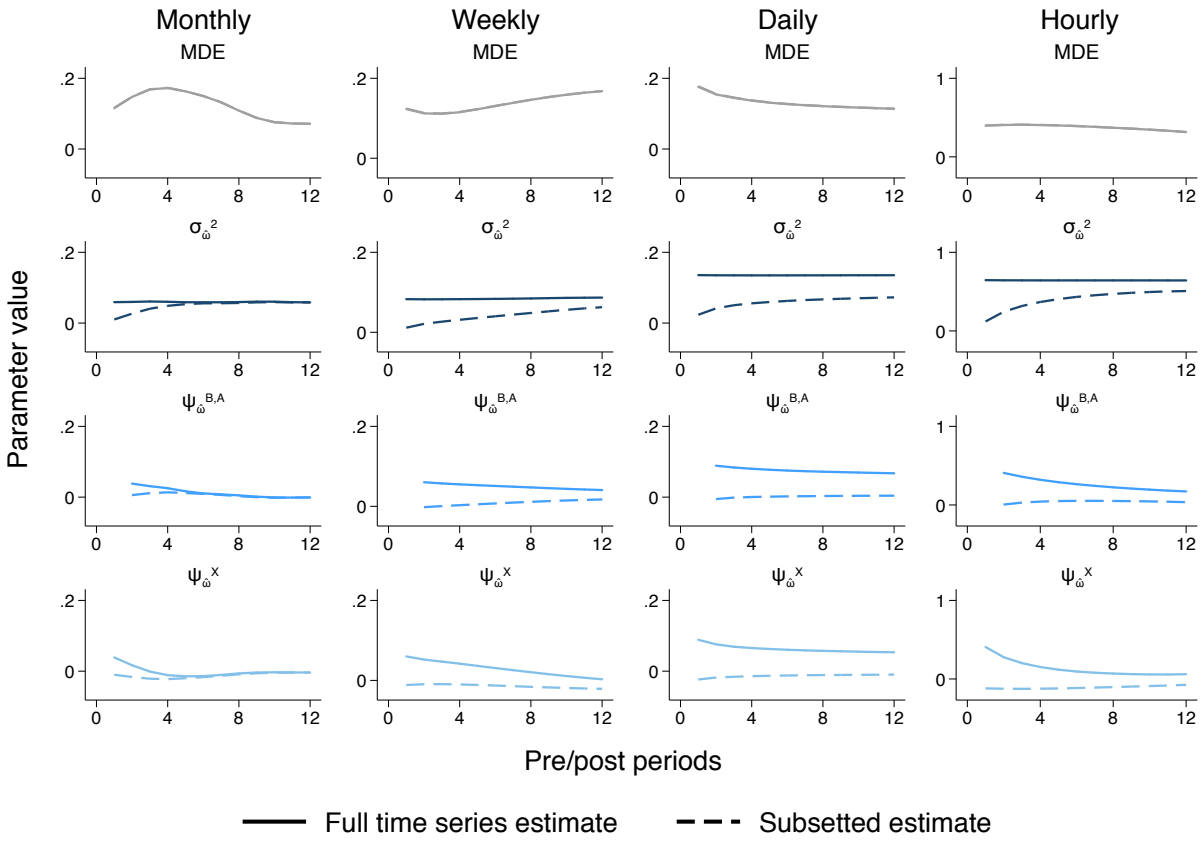
[26]The STATA program `sampsi` assumes a normal distribution, which is not appropriate for small samples.

Figure B.7: Estimated parameters – Bloom et al. (2015) data



Notes: This figure shows two different methods for estimating variance parameters, applied to the Bloom et al. (2015) data, and also depicts the resulting minimum detectable effect (MDE) resulting from both methods. The solid lines plot parameters calculated by running a regression of the outcome variable on unit and time period fixed effects, estimated on the entire time series of data, and then using the residuals from this regression to calculate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ for the average panel of length $m + r$. We then plug these estimates into Equation (3.2) to calculate the minimum detectable effect. The dashed lines show parameters estimated using the procedure described above, where rather than use residuals from the full time series, we subset the dataset into shorter panels of length $m + r$, calculate the parameters using residuals only from this subset, and average across all possible subsets to arrive at $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. We calculate the MDE by plugging these estimates into Equation (B.86). Note that these variance-covariance estimates converge as the panel length increases. Both procedures yield (virtually) identical MDEs, even though the underlying parameter estimates differ substantially.

Figure B.8: Estimated parameters – Pecan Street data



Notes: This figure shows two different methods for estimating variance parameters, applied to the Pecan Street data at the four levels of aggregation presented in the main text, and also depicts the resulting minimum detectable effect (MDE) resulting from both methods. Note that the y axis scale differs between the hourly data and the other three datasets; this is because the degree of residual variation left in the hourly data after removing time and individual fixed effects is much greater than in the other datasets. The solid lines show parameters calculated by running a regression of the outcome variable on unit and time period fixed effects, estimated on the entire time series of data, and then using the residuals from this regression to calculate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ for the average panel of length $m + r$. We then plug these estimates into Equation (3.2) to calculate the minimum detectable effect. By contrast, the dashed lines show parameters estimated using the procedure described above, where rather than use residuals from the full time series, we subset the dataset into shorter panels of length $m + r$, calculate the parameters using residuals only from this subset, and average across all possible subsets to arrive at $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. We calculate the MDE by plugging these estimates into Equation (B.86). Note that these variance-covariance estimates converge as the panel length increases. Both procedures yield (virtually) identical MDEs, even though the two procedures' method for estimating the underlying parameters differ substantially.

B.4.2 Simulation-based power calculations

In cases where researchers have access to a representative pre-existing dataset, we recommend that they perform power calculations via simulation. This obviates the need to estimate *ex ante* variance-covariance parameters, and it ensures that *ex ante* power calculations assume the same experimental design, regression model, and variance estimator expected to be used in *ex post* analysis. Our accompanying STATA package facilitates simulation-based power calculations using the program `pc_simulate`. This program implements the following algorithm:

1. Choose the following candidate parameters: sample size J, pre-treatment periods m, and post-treatment periods r, treatment ratio P, minimum detectable effect MDE, and significance level α. Let X_{it} denote the outcome variable of interest in the pre-existing dataset.

2. Randomly select J units from the representative dataset, and randomly select a range of $(m+r)$ consecutive time periods. This will serve as a simulated experimental dataset, with sample size J, m pre-treatment periods, and r post-treatment periods.

3. Randomly scramble a $[J \times 1]$ vector of PJ ones and $(1-P)J$ zeros, rounding PJ to the nearest integer. Assign each of the J units to either treatment $(D = 1)$ or control $(D = 0)$, based on the order of this scrambled vector.

4. Construct an experimental outcome variable Y_{it}, where $Y_{it} = X_{it} + MDE$ for treated units in post-treatment periods, and $Y_{it} = X_{it}$ otherwise.

5. Using this simulated experimental dataset and the simulated outcome variable Y_{it}, implement the exact regression specification and variance estimator to be used in *ex post* analysis. Record whether this model rejects the null hypotheses of zero treatment effects with significance level α (i.e. $H_0 : \tau = 0$).

6. Repeat Steps 2–5 many times, and calculate the rejection rate across all simulations. This is the experiment's statistical power as a function of J, m, r, P, MDE, and α.

7. Repeat Steps 1–6 for a range of MDEs and design parameters, increasing the number of simulations after narrowing down this range of parameters to more precisely calibrate power.

This algorithm allows users to test alternative regression specifications and alternative standard error assumptions, without needing to formally derive a power calculation expression for each model. If the pre-existing dataset contains fewer cross-sectional units than the desired sample size J, `pc_simulate` lets users simulate additional units by bootstrapping units with replacement from the existing dataset. Unfortunately, if the pre-existing dataset contains fewer time periods than the desired panel length $(m + r)$, an analogous bootstrapping procedure would be much less straightforward (because unlike cross-sectional units, time periods are ordered and have a ordered covariance structure that is not orthogonal to the treatment vector D).

This simulation-based algorithm can only calibrate statistical power κ. Rather than rely on the critical value $t^d_{1-\kappa}$, the algorithm simply estimates realized power as the proportion of simulations where the treatment effect is statistically distinguishable from zero. (By contrast, users may algebraically rearrange (or invert) an analytical power calculation formula to solve for any one of its parameters.) Calibrating simulation-based power calculations for a parameter other than κ necessitates a grid search over candidate parameter values, as described in Step 7 above. For example, to calibrate sample size J by simulation, users may repeat Step 1–6 over a range of candidate J values, narrowing this range (while simultaneously increasing the number of simulations) to calibrate to the desired power.

B.4.3 Lack of (representative) pre-existing data

To perform accurate *ex ante* power calculations, researchers must either have access to data that is representative (in expectation) of their future experimental data, or be able to parameterize an analytical formula with accurate estimates of the variance and covariance of the error structure. We recommend that simulation-based power calculations (as described above) in cases where researchers have a representative pre-existing dataset with (i) data for the desired outcome (and relevant control variables); (ii) at least as many unique cross-sectional units as the desired experimental sample size; and (iii) a time series at least as long as the desired experimental panel length. Many candidate experiments likely satisfy these criteria, such as when researchers partner with organizations that maintain historical databases on the desired population of experimental subjects.

At the same time, researchers may lack access to representative data *ex ante*. This problem is not unique to panel data, as even the simple cross-sectional power calculation formula (Equation (B.6)) hinges on (an estimate of) the variance σ^2_ε. However, power calculations for panel RCT designs require four variance-covariance parameters: σ^2_ω, ψ^B, ψ^A, and ψ^X. While σ^2_ω is fixed in the population, the ψ (and $\psi_{\hat{\omega}}$) terms are endogenous to the panel length of the experiment, which underscores the importance of estimating $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from a representative time series.

In the absence of representative data, researchers may using analytical formulas in conjunction with appropriate sensitivity analyses.[27] Depending on the type of data that *is* available, approximating the parameters σ^2_ω, ψ^B, ψ^A, and ψ^X may be possible. We consider four cases:

1. *Too few units:* If researchers have access to a representative pre-existing dataset with too few cross-sectional units, they may still estimate $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, and apply these values to the (estimation-specific) analytic formula.[28] These variance-covariance parameters do *not* depend on sample size J in the SCR power calculation formula, and estimates of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ derived from residuals are *not* sensitive in expecta-

[27] An alternative is to either impose assumptions on some existing dataset or construct a simulated dataset, either of which could be used to conduct power calculations by simulation. However, this process will be much more computationally intensive than simply applying an analytical formula with appropriate parameter sensitivities.

[28] Researchers using *estimated* parameters should apply Equation (B.86) rather than Equation (3.2).

tion to the number of panel units J to be used in the experiment.[29] Alternatively, we recommend that researchers bootstrap units by sampling existing units with replacement, and use this expanded dataset (including simulated units) to conduct power calculations by simulation, as described above.[30]

2. *Too few time periods:* If the pre-existing dataset contains too few time periods, researchers may still estimate $\tilde{\sigma}^2_{\hat{\omega}}$ using residuals from a regression with fewer than $m+r$ periods (because σ^2_{ω} does not depend on panel length). However, the ψ terms do depend on panel length, and they cannot be estimated directly from a dataset with fewer than $m+r$ periods. One strategy is to simply estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ using the longest possible panel (i.e., all available time periods in the pre-existing dataset), even if it is shorter than $m+r$ periods. The resulting $\psi_{\hat{\omega}}$ estimates are likely to be upper bounds (in absolute value) on the $\psi_{\hat{\omega}}$ estimates for longer panels, because as the panel length increases, the ψ terms incorporate more covariances between time periods that are further apart (which tend to become less correlated in distance). Another strategy is to attempt to extend the time series for each unit, analogous to the approach of bootstrapping units. As a rule of thumb, researchers often approximate time series data as an AR(k) process with $k \geq \sqrt[3]{T}$, where T is the full time series length. To extend short panels, researchers may estimate this AR(k) process using (residuals from) the existing dataset, and then simulate forward for each unit's outcome realization. Neither of these strategies is perfect, and we recommend conducting appropriate sensitivity analysis in either case.

3. *No data, standard cross-sectional or DD model:* In the complete absence of data, power calculations will be challenging. At the very least, we recommend that researchers search for estimates of the residual variance in the existing literature, noting that panel fixed effects models are likely to yield lower residual variances than cross-sectional models with similar outcome data. If this is not possible, researchers may iterate analytical power calculations over a range of parameter choices. If researchers are able to guess a reasonable value of σ^2_{ω}, they my test a range of AR(1) parameters for plausible values of $\psi^{B,A,X}$. As a rule of thumb, $\psi^{B,A,X}$ are likely to be positive in the absence of a strong prior of negative serial correlation. In absolute value, $\psi^{B,A,X}$ should not exceed σ^2_{ω}, and they should decrease monotonically in panel length. To provide a sense of what plausible values of ψ^B, ψ^A, and ψ^X (and their residual-based counterparts) may be, we plot estimates from a range of panel lengths using simulated AR(1) data, the Bloom et al. (2015) data, and Pecan Street data, in Figures B.6, B.7, and B.8, respectively.

4. *No data, other models:* For more complicated RCT designs (or ANCOVA in the presence of time shocks), analytical power calculation formulas might not align with the

[29] Estimates of $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ are sensitive to the number of cross-sectional units I used to estimate $\hat{\omega}_{it}$, but this is not related to the sample size parameter J. In Equation (B.86), the comparative static $d\psi_{\hat{\omega}}/dJ = 0$.

[30] The `bootstrap` option in `pc_simulate` does exactly this, drawing units with replacement until reaching the desire sample size.

desired *ex post* estimating equation. This would render our above Case 3 recommendation inactionable. An alternative strategy (in the absence of data) is to simulate data based on a simple DGP, incorporating AR(1) serial correlation into the idiosyncratic error term. Researchers can perform simulation-based power calculations using these simulated data, iterating over multiple simulated datasets to test for sensitivity to DGP (and AR(1)) assumptions.

B.4.4 STATA package `pcpanel`

To facilitate user implementation of the methods described above, we have built an accompanying STATA package `pcpanel`. This package contains three programs, which we introduce iin Section 3.4.2 of the main text. We describe each function in detail below:

`pc_dd_analytic`

This program performs analytical power calculations using our SCR formula for difference-in-differences (Equation (3.2)). Users may input any two of the options {`mde` (effect size MDE), `n` (sample size J), `power` (κ)}, and the program will solve Equation (3.2) for the third option. Users may also adjust four additional experimental design options: treatment ratio `p` (P), number of pre-treatment periods `pre` (m), number of post-treatment periods `post` (r), and false rejection rate `alpha` (α). `pc_dd_analytic` parameterizes the variance (σ^2_ω) and covariance (ψ^B, ψ^A, ψ^X) in Equaton (3.2) in one of two ways.

First, users may allow the subprogram `pc_dd_covar` to nonparametrically estimate the idiosyncratic residual variance/covariance papers. To do this, they input the outcome variable of interest from the dataset in memory (using option `depvar`), as well as variables that identify units and time periods (using options `i` and `t`, respectively). This is our preferred means of parameterizing Equation (3.2) in practice, using pre-existing data.

Second, users may manually input *either* an idiosyncratic residual variance (option `variance` (σ^2_ω)) *or* an idiosyncratic residual standard deviation (option `sd` (σ_ω)). They may also incorporate non-i.i.d. idiosyncratic errors in 1 of 3 ways: (1) input assumed AR(1) parameter(s) (option `ar1` (γ)) which `pc_dd_analytic` uses to calculate average covariances (ψ^B, ψ^A, ψ^X); (2) input assumed average covariances themselves (option `avgcov`); or (3) input assumed average correlations (option `avgcor`), which the program multiplies by the variance to convert to covariances. Specifying `var` (or `sd`) without specifying `ar1`, `avgcov`, or `avgcor` will cause `pc_dd_analytic` to default to assuming $\psi^B = \psi^A = \psi^X = 0$.[31] For users accustomed to `sampsi`, which accepts the *composite* residual standard deviation and intracluster correlation (σ_ε and ρ, in the notation of McKenzie (2012)), the `pc_dd_analytic` option `var` combines these two terms into a single input (σ^2_ω, in our notation) where $\sigma^2_\omega = \sigma^2_\varepsilon(1-\rho)$.[32]

[31] Given that estimating ψ^B, ψ^A, ψ^X is nontrivial (see Appendix B.4.1), users should apply the options `avgcov` and `avgcor` with caution.

[32] In the absence of time shocks (i.e. $\sigma^2_\delta = 0$, as in McKenzie (2012)), this equality holds. However, if the data generating process includes time shocks, then $\sigma^2_\omega = \sigma^2_\varepsilon(1-\rho_\upsilon-\rho_\delta)$, where ρ_υ is the intracluster correlation within units and ρ_δ is the intracluster correlation within time periods.

`pc_dd_covar`

This program nonparametrically estimates $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ from an existing dataset (i.e. the dataset in memory), for a given number of pre-treatment and post-treatment periods. Users must input the outcome variable in question, as well as the number or pre-treatment periods (option `pre` (m)), post-treatment periods (opton `post` (r)), unit identifier (option `i`) and time period identifier (option `t`). Given these inputs, `pc_dd_covar` operationalizes the procedure outlined in Appendix B.4.1. Most users will not need to directly apply this program, as it is designed to operate "under the hood" of `pc_dd_analytic`.

`pc_simulate`

This program performs power calculations by simulation, following the algorhtim in Appendix B.4.2. it supports four types of RCTs:

- `ONESHOT` (one wave of post-treatment data)
- `POST` (multiple waves of post-treatment data)
- `DD` (pre-treatment and post-treatment data)
- `ANCOVA` (post-treatment data, conditioning on pre-treatment data)

Users may input six basic experimental design options: effect size `mde` (MDE), sample size `n` (J), treatment ratio `p` (P), number of pre-treatment periods `pre` (m), number of post-treatment periods `post` (r), and false rejection rate `alpha` (α). Given these options, `pc_simulate` calculates the average rejection rate (i.e. power (κ)) of the treatment effect estimator, across all simulations.

The program is sufficiently flexible to allow for linear controls (option `controls`), fixed effects (options `absorb`, `absorbfactor`), regression weights (opton `weight`) and standard errors (option `vce`). Each of these options, when specified, is passed through the regression function `reghdfe` in each simulation.[33] Toggling the option `bootstrap` causes the program to sample J units with replacement for each simulation, which facilitates power calculations with a greater number of units than are present in the dataset in memory. If `bootstrap` is not toggled, each simulation samples J units without replacement. Users may also restrict the range of time periods over which simulations occur (option `tstart`), and the number of simulations for each set of parameters (option `nsim`).[34]

Instead of applying the CRVE, Bertrand, Duflo, and Mullainathan (2004) also recommend collapsing panel datasets to a single pre-treatment and post-treatment observation. `pc_simulate` lets users conduct power calculations on collapsed panel estimators, by toggling the option `collapse`. For `POST` and `ANCOVA` models, each simulation collapses the estimating

[33] `pc_simulate` simulates regressions using the function `reghdfe` (`http://scorreia.com/software/reghdfe/`).

[34] The program defaults users to 500 simulations. Increasing the number of simulations will improve accuracy at the expense of runtime.

equation to a cross-sectional model with one observation per unit (where the dependent variable is averaged within units across all r time periods). For the `DD` model, each simulation collapses the estimating equation to a two-period panel, with one pre-treatment and one post-treatment observation per unit (where the dependent variable is averaged within units across all m or r time periods).

`pc_simulate` supports stratified randomization via the option `stratify`. When users specify one or more categorical variables (e.g. `stratify(gender race)`), the program separately randomizes P units within each `gender-race` cell. For stratified randomization, the option `n` governs the number of units within each *cell* (as opposed to the total number of units in the full sample). The program continues to estimate a single *pooled* average treatment effect. If researchers are interested in heterogeneous treatment effects by cell, they can iterate `pc_simulate` separately for each stratification cell—leaving `stratify` unspecified, and conditioning on stratification cells using the option `if` (e.g. `pc_simulate y if gender=="male" & race=="white", ...`).

`pc_simulate` also supports cluster randomization via the option `idcluster`. When users specify a group identifier (e.g. `idcluster(village_id)`), the program randomizes at the group level, where whole groups are randomly assigned to treatment/control. For cluster randomization, the option `n` governs the number of *groups* (e.g. villages), as opposed to the total number of individual units (e.g. households). The program continues to estimate a single *pooled* average treatment effect at the unit level. `pc_simulate` also supports two cluster randomization suboptions. The suboption `sizecluster` governs the number of units sampled within each cluster; if not specified, the program defaults to the size of each cluster in the existing dataset. The suboption `pcluster` governs the intensity of treatment within treated clusters: (i) if `pcluster` is not specified, the program defaults to a treatment intensity of 1 in all treated clusters (i.e. all units in treated clusters receive treatment); (ii) if `pcluster` is specified with a single value (e.g. `pcluster(0.5)`), the program randomly assigns treatment to this proportion (i.e. 50 percent) of units within each treated cluster (all units in control clusters remain untreated); (iii) if `pcluster` is specified with multiple values (e.g. `pcluster(0.33 0.67 1)`), the program varies the intensity of treatment in equal proportions across treated clusters (i.e. one third of treated clusters receive `pcluster(0.33)`, etc.).

B.5 Estimation-related proofs

In this section, we prove that researchers may calculate unbiased power calculations by estimating the variance-covariance parameters from a real dataset, where the parameters governing the data generating process is unknown.

B.5.1 Recovering estimated parameters

Here, we demonstrate that the procedure described in Appendix B.4.1 recovers unbiased estimates of the variance and covariance parameters governing the residuals $\hat{\omega}_{it}$ from a regression of Y_{it} on unit and time fixed effects, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ (though these do *not*

represent unbiased estimates of the *true* parameters σ^2_ω, ψ^B, ψ^A, and ψ^X). We denote our procedure for computing these parameters with a $\sim$. Note that throughout this section, we are considering I units in the sample used to estimate $\hat{\omega}_{it}$, which may be distinct from the sample size J units used in the ensuing power calculations. Note also that because we are estimating the variance and covariance of a *population* of residuals, we use the population variance/covariance estimators as opposed to the (unbiased) sample variance/covariance estimators.[35]

In order to estimate the variance of the residuals, $\sigma^2_{\hat{\omega}}$, we define:

$$\tilde{\sigma}^2_{\hat{\omega}} \equiv \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \left(\hat{\omega}_{it} - \overline{\overline{\hat{\omega}}} \right)^2$$

where $\overline{\overline{\hat{\omega}}} = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \hat{\omega}_{it} = 0$. Taking expectations of both sides:

$$\begin{aligned} \mathbb{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X} \right] &= \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \mathbb{E}\left[\hat{\omega}^2_{it} \mid \mathbf{X} \right] \\ &= \sigma^2_{\hat{\omega}} \end{aligned}$$

To estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, we define the $[I \times 1]$ vector of residuals for period t as $\hat{\boldsymbol{\omega}}_t$ and calculate the average covariance between any two vectors in the relevant range of time periods. For the pre-period, we define:

$$\begin{aligned} \tilde{\psi}^B_{\hat{\omega}} &\equiv \frac{2}{Im(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \sum_{i=1}^{I} \left(\hat{\omega}_{it} - \overline{\hat{\omega}}_t \right) \left(\hat{\omega}_{is} - \overline{\hat{\omega}}_s \right) \\ &= \frac{2}{Im(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \sum_{i=1}^{I} \left(\hat{\omega}_{it}\hat{\omega}_{is} - \hat{\omega}_{it}\overline{\hat{\omega}}_s - \hat{\omega}_{is}\overline{\hat{\omega}}_t + \overline{\hat{\omega}}_t\overline{\hat{\omega}}_s \right) \end{aligned}$$

where $\overline{\hat{\omega}}_t = \frac{1}{I} \sum_{i=1}^{I} \hat{\omega}_{it} = 0$. Taking expectations yields:

$$\begin{aligned} \mathbb{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X} \right] &= \frac{2}{Im(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \sum_{i=1}^{I} \mathbb{E}\left[\hat{\omega}_{it}\hat{\omega}_{is} \mid \mathbf{X} \right] \\ &= \frac{2}{Im(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \sum_{i=1}^{I} \mathrm{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \end{aligned}$$

[35] This means that to calculate $\tilde{\sigma}^2_{\hat{\omega}}$, we deflate the sample variance estimate by $\frac{IT-1}{IT}$, and to calculate the $\tilde{\psi}_{\hat{\omega}}$ terms, we deflate the sample covariance estimates by $\frac{I-1}{I}$. This distinction is ultimately innocuous, and the following derivations simply rely on a consistent decision to use either the population or sample variance/covariance estimators.

$$= \psi^B_{\hat{\omega}}$$

Similarly:

$$\begin{aligned}\tilde{\psi}^A_{\hat{\omega}} &\equiv \frac{2}{Ir(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \sum_{i=1}^{I} \left(\hat{\omega}_{it} - \overline{\hat{\omega}}_t\right)\left(\hat{\omega}_{is} - \overline{\hat{\omega}}_s\right) \\ &= \frac{2}{Ir(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \sum_{i=1}^{I} \left(\hat{\omega}_{it}\hat{\omega}_{is} - \hat{\omega}_{it}\overline{\hat{\omega}}_s - \hat{\omega}_{is}\overline{\hat{\omega}}_t + \overline{\hat{\omega}}_t\overline{\hat{\omega}}_s\right)\end{aligned}$$

and therefore:

$$\mathbb{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right] = \psi^A_{\hat{\omega}}$$

Applying the same steps to $\psi^X_{\hat{\omega}}$:

$$\begin{aligned}\tilde{\psi}^X_{\hat{\omega}} &\equiv \frac{1}{Imr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \sum_{i=1}^{I} \left(\hat{\omega}_{it} - \overline{\hat{\omega}}_t\right)\left(\hat{\omega}_{is} - \overline{\hat{\omega}}_s\right) \\ &= \frac{1}{Imr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \sum_{i=1}^{I} \left(\hat{\omega}_{it}\hat{\omega}_{is} - \hat{\omega}_{it}\overline{\hat{\omega}}_s - \hat{\omega}_{is}\overline{\hat{\omega}}_t + \overline{\hat{\omega}}_t\overline{\hat{\omega}}_s\right)\end{aligned}$$

Taking expectations of both sides:

$$\begin{aligned}\mathbb{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right] &= \frac{1}{Imr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \sum_{i=1}^{I} \mathbb{E}\left[\hat{\omega}_{it}\hat{\omega}_{is} \mid \mathbf{X}\right] \\ &= \frac{1}{Imr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \sum_{i=1}^{I} \mathrm{Cov}\left(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}\right) \\ &= \psi^X_{\hat{\omega}}\end{aligned}$$

Hence, we can recover unbiased estimates of the parameters $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ (defined over residuals $\hat{\omega}_{it}$, rather than errors ω_{it}) by calculating the averages of the estimated $\tilde{\sigma}^2_{\hat{\omega}}$, $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$, respectively.

B.5.2 Estimating MDE from residual-based parameters

To calculate the MDE using the SCR formula, we must know the true parameters that characterize the variance and covariance of the error structure, σ^2_{ω}, ψ^B, ψ^A, and ψ^X. We cannot calculate these parameters directly from a real dataset, however, because we do not observe the true error structure or data generating process. Instead, we estimate a residual for each observation and calculate the residual-based analogs of these parameters, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. In this section, we derive an expression for MDE^{est} in terms of these residual-based parameters that is equivalent to MDE from the SCR formula as defined in terms of true variance-covariance parameters:

$$MDE^{est}\left(\sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}\right) = MDE\left(\sigma^2_{\omega}, \psi^B, \psi^A, \psi^X\right)$$

Model While estimating the variance and covariance parameters of a dataset does not require a treatment, we assume that all other features of this model are identical to the model that generates the serial-correlation-robust power calculation formula, Equation (3.2).

That is, there are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods. We restate the remaining assumptions from Appendix B.1.2.2 here for convenience:

Assumption (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \delta_t + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a unit-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; δ_t is a time-specific disturbance distributed i.i.d. $\mathcal{N}(0, \sigma_\delta^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*,* $J - 1$ *unit dummies, and* $(m + r) - 1$ *time dummies. This again follows from random assignment of* D_{it}*.*

Assumption (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption (Independence across units)**.** $\mathbb{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ \forall\ t, s.$

Assumption (Symmetric covariance structures)**.** *Define:*

$$\psi^B \equiv \frac{2}{Jm(m-1)} \sum_{i=1}^{J} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^A \equiv \frac{2}{Jr(r-1)} \sum_{i=1}^{J} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi^X \equiv \frac{1}{Jmr} \sum_{i=1}^{J} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of the same unit, respectively. Define ψ_T^B, ψ_T^A, and ψ_C^X analogously, where we consider only the PJ treated units; also define ψ_C^B, ψ_C^A, and ψ_C^X analogously, where we consider only the $(1 - P)J$ control units. Using these definitions, assume that $\psi^B = \psi_T^B = \psi_C^B$; $\psi^A = \psi_T^A = \psi_C^A$; and $\psi^X = \psi_T^X = \psi_C^X$.

Parameters We first need to estimate the residuals of this model. To do this, we regress Y_{it} on a constant and fixed effects at the unit and time levels. For a balanced panel, the estimated coefficients are

$$\begin{aligned}
\hat{\beta} &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} Y_{it} \\
\hat{\upsilon}_i &= \frac{1}{T}\sum_{t=-m+1}^{r} Y_{it} - \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} Y_{it} \\
\hat{\delta}_t &= \frac{1}{I}\sum_{i=1}^{I} Y_{it} - \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} Y_{it}
\end{aligned}$$

Then the residual is defined as

$$\begin{aligned}
\hat{\omega}_{it} &= Y_{it} - \hat{Y}_{it} \\
&= (\beta + \upsilon_i + \delta_t + \omega_{it}) - \left(\hat{\beta} + \hat{\upsilon}_i + \hat{\delta}_t\right) \\
&= \omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}}
\end{aligned}$$

where

$$\begin{aligned}
\overline{\omega}_i &= \frac{1}{T}\sum_{t=-m+1}^{r} \omega_{it} \\
\overline{\omega}_t &= \frac{1}{I}\sum_{i=1}^{I} \omega_{it} \\
\overline{\overline{\omega}} &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \omega_{it}
\end{aligned}$$

We can now use this definition of residuals to derive expressions for $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. We first derive an expression for $\sigma^2_{\hat{\omega}}$, the average variance of a residual:

$$\begin{aligned}
\sigma^2_{\hat{\omega}} &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}(\hat{\omega}_{it} \mid \mathbf{X}) \\
&= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \operatorname{Var}\left(\omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} \Big[\operatorname{Var}(\omega_{it} \mid \mathbf{X}) + \operatorname{Var}\left(\overline{\omega}_i \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{\omega}_t \mid \mathbf{X}\right) + \operatorname{Var}\left(\overline{\overline{\omega}} \mid \mathbf{X}\right) \\
&\qquad - 2\operatorname{Cov}\left(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\omega_{it}, \overline{\omega}_t \mid \mathbf{X}\right) + 2\operatorname{Cov}\left(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}\right) \\
&\qquad + 2\operatorname{Cov}\left(\overline{\omega}_i, \overline{\omega}_t \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}\right) - 2\operatorname{Cov}\left(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}\right) \Big]
\end{aligned}$$

Calculating each of these terms, in turn, gives

$$\begin{aligned}
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}(\omega_{it}\mid\mathbf{X}) &= \sigma_\omega^2\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}(\overline{\omega}_i\mid\mathbf{X}) &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}\left(\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is}\mid\mathbf{X}\right)\\
&= \frac{1}{T}\sigma_\omega^2+\frac{m(m-1)}{T^2}\psi^B+\frac{r(r-1)}{T^2}\psi^A+\frac{2mr}{T^2}\psi^X\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}(\overline{\omega}_t\mid\mathbf{X}) &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}\mid\mathbf{X}\right)\\
&= \frac{1}{I}\sigma_\omega^2\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}(\overline{\overline{\omega}}\mid\mathbf{X}) &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Var}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{s=-m+1}^{r}\omega_{js}\mid\mathbf{X}\right)\\
&= \frac{1}{IT}\sigma_\omega^2+\frac{m(m-1)}{IT^2}\psi^B+\frac{r(r-1)}{IT^2}\psi^A+\frac{2mr}{IT^2}\psi^X\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}2\operatorname{Cov}(\omega_{it},\overline{\omega}_i\mid\mathbf{X}) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Cov}\left(\omega_{it},\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is}\mid\mathbf{X}\right)\\
&= \frac{2}{T}\sigma_\omega^2+\frac{2m(m-1)}{T^2}\psi^B+\frac{2r(r-1)}{T^2}\psi^A+\frac{4mr}{T^2}\psi^X\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}2\operatorname{Cov}(\omega_{it},\overline{\omega}_t\mid\mathbf{X}) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Cov}\left(\omega_{it},\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}\mid\mathbf{X}\right)\\
&= \frac{2}{I}\sigma_\omega^2\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}2\operatorname{Cov}(\omega_{it},\overline{\overline{\omega}}\mid\mathbf{X}) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Cov}\left(\omega_{it},\frac{1}{IT}\sum_{j=1}^{I}\sum_{s=-m+1}^{r}\omega_{js}\mid\mathbf{X}\right)\\
&= \frac{2}{IT}\sigma_\omega^2+\frac{2m(m-1)}{IT^2}\psi^B+\frac{2r(r-1)}{IT^2}\psi^A+\frac{4mr}{IT^2}\psi^X\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}2\operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_t\mid\mathbf{X}) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is},\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}\mid\mathbf{X}\right)\\
&= \frac{2}{IT}\sigma_\omega^2+\frac{2m(m-1)}{IT^2}\psi^B+\frac{2r(r-1)}{IT^2}\psi^A+\frac{4mr}{IT^2}\psi^X\\
\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}2\operatorname{Cov}(\overline{\omega}_i,\overline{\overline{\omega}}\mid\mathbf{X}) &= \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{s=-m+1}^{r}\omega_{is},\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}\mid\mathbf{X}\right)\\
&= \frac{2}{IT}\sigma_\omega^2+\frac{2m(m-1)}{IT^2}\psi^B+\frac{2r(r-1)}{IT^2}\psi^A+\frac{4mr}{IT^2}\psi^X
\end{aligned}$$

$$\frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r} 2\operatorname{Cov}\left(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}\right) = \frac{2}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}, \frac{1}{IT}\sum_{k=1}^{I}\sum_{s=-m+1}^{r}\omega_{ks} \mid \mathbf{X}\right)$$
$$= \frac{2}{IT}\sigma_\omega^2 + \frac{2m(m-1)}{IT^2}\psi^B + \frac{2r(r-1)}{IT^2}\psi^A + \frac{4mr}{IT^2}\psi^X$$

Combining these terms and simplifying yields

$$\sigma_{\hat{\omega}}^2 = \left(\frac{(I-1)(T-1)}{IT}\right)\sigma_\omega^2 - \left(\frac{(I-1)m(m-1)}{IT^2}\right)\psi^B - \left(\frac{(I-1)r(r-1)}{IT^2}\right)\psi^A - \left(\frac{2(I-1)mr}{IT^2}\right)\psi^X \tag{B.87}$$

We next derive an expression for $\psi_{\hat{\omega}}^A$:

$$\begin{aligned}
\psi_{\hat{\omega}}^A &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}}, \omega_{is} - \overline{\omega}_i - \overline{\omega}_s + \overline{\overline{\omega}} \mid \mathbf{X}) \\
&= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\Big[\operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \operatorname{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) - \operatorname{Cov}(\omega_{it}, \overline{\omega}_s \mid \mathbf{X}) + \operatorname{Cov}(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad - \operatorname{Cov}(\overline{\omega}_i, \omega_{is} \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i, \overline{\omega}_i \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i, \overline{\omega}_s \mid \mathbf{X}) - \operatorname{Cov}(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad - \operatorname{Cov}(\overline{\omega}_t, \omega_{is} \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t, \overline{\omega}_i \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t, \overline{\omega}_s \mid \mathbf{X}) - \operatorname{Cov}(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad + \operatorname{Cov}(\overline{\overline{\omega}}, \omega_{is} \mid \mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\omega}_i \mid \mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\omega}_s \mid \mathbf{X}) + \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\overline{\omega}} \mid \mathbf{X})\Big]
\end{aligned}$$

We again calculate each term:

$$\begin{aligned}
\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) &= \psi^A \\
\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it}, \frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip} \mid \mathbf{X}\right) \\
&= \frac{2}{r(r-1)IT}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it}, \omega_{ip} \mid \mathbf{X}) \\
\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \overline{\omega}_s \mid \mathbf{X}) &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it}, \frac{1}{I}\sum_{j=1}^{I}\omega_{js} \mid \mathbf{X}\right) \\
&= \frac{1}{I}\psi^A \\
\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}) &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\omega_{it}, \frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp} \mid \mathbf{X}\right) \\
&= \frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it}, \omega_{ip} \mid \mathbf{X})
\end{aligned}$$

$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\omega_{is}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\omega_{is}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)IT}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}\left(\omega_{it},\omega_{ip}\mid\mathbf{X}\right)$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_i\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq}\mid\mathbf{X}\right)$$
$$=\frac{1}{T}\sigma_\omega^2+\frac{m(m-1)}{T^2}\psi^B+\frac{r(r-1)}{T^2}\psi^A+\frac{2mr}{T^2}\psi^X$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\overline{\omega}_s\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{I}\sum_{j=1}^{I}\omega_{js}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_i,\overline{\overline{\omega}}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip},\frac{1}{IT}\sum_{j=1}^{I}\sum_{q=-m+1}^{r}\omega_{jq}\mid\mathbf{X}\right)$$
$$=\frac{1}{IT}\sigma_\omega^2+\frac{m(m-1)}{IT^2}\psi^B+\frac{r(r-1)}{IT^2}\psi^A+\frac{2mr}{IT^2}\psi^X$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\omega_{is}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\omega_{is}\mid\mathbf{X}\right)$$
$$=\frac{1}{I}\psi^A$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\overline{\omega}_i\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\overline{\omega}_s\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{I}\sum_{k=1}^{I}\omega_{ks}\mid\mathbf{X}\right)$$
$$=\frac{1}{I}\psi^A$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\omega}_t,\overline{\overline{\omega}}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt},\frac{1}{IT}\sum_{k=1}^{I}\sum_{p=-m+1}^{r}\omega_{kp}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(r-t)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\omega_{is}\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\omega_{is}\mid\mathbf{X}\right)$$
$$=\frac{2}{r(r-1)I^2T}\sum_{i=1}^{I}\sum_{t=1}^{r}\sum_{p=-m+1}^{r}(t-1)\operatorname{Cov}(\omega_{it},\omega_{ip}\mid\mathbf{X})$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\overline{\omega}_i\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq}\mid\mathbf{X}\right)$$
$$=\frac{1}{IT}\sigma_\omega^2+\frac{m(m-1)}{IT^2}\psi^B+\frac{r(r-1)}{IT^2}\psi^A+\frac{2mr}{IT^2}\psi^X$$
$$\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}(\overline{\overline{\omega}},\overline{\omega}_s\mid\mathbf{X})=\frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\operatorname{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp},\frac{1}{I}\sum_{k=1}^{I}\omega_{ks}\mid\mathbf{X}\right)$$

$$= \frac{2}{r(r-1)I^2T} \sum_{i=1}^{I} \sum_{t=1}^{r} \sum_{p=-m+1}^{r} (t-1) \operatorname{Cov}(\omega_{it}, \omega_{ip} \mid \mathbf{X})$$

$$\frac{2}{r(r-1)I} \sum_{i=1}^{I} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \operatorname{Cov}(\overline{\omega}, \overline{\omega} \mid \mathbf{X}) = \frac{2}{r(r-1)I} \sum_{i=1}^{I} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \operatorname{Cov}\left(\frac{1}{IT} \sum_{j=1}^{I} \sum_{p=-m+1}^{r} \omega_{jp}, \frac{1}{IT} \sum_{k=1}^{I} \sum_{q=m+1}^{r} \omega_{kq} \mid \mathbf{X} \right)$$

$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X$$

Combining these terms and simplifying yields

$$\psi_{\hat{\omega}}^A = -\left(\frac{I-1}{IT}\right)\sigma_\omega^2 + \left(\frac{(I-1)m(m-1)}{IT^2}\right)\psi^B + \left(\frac{(I-1)(m^2+2m+r)}{IT^2}\right)\psi^A - \left(\frac{2(I-1)m^2}{IT^2}\right)\psi^X \tag{B.88}$$

By symmetry, the expression for $\psi_{\hat{\omega}}^B$ is

$$\psi_{\hat{\omega}}^B = -\left(\frac{I-1}{IT}\right)\sigma_\omega^2 + \left(\frac{(I-1)(r^2+2r+m)}{IT^2}\right)\psi^B + \left(\frac{(I-1)r(r-1)}{IT^2}\right)\psi^A - \left(\frac{2(I-1)r^2}{IT^2}\right)\psi^X \tag{B.89}$$

We finally derive an expression for $\psi_{\hat{\omega}}^X$:

$$\begin{aligned}
\psi_{\hat{\omega}}^X &= \frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}(\omega_{it} - \overline{\omega}_i - \overline{\omega}_t + \overline{\overline{\omega}}, \omega_{is} - \overline{\omega}_i - \overline{\omega}_s + \overline{\overline{\omega}} \mid \mathbf{X}) \\
&= \frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \Big[\operatorname{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \operatorname{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) - \operatorname{Cov}(\omega_{it}, \overline{\omega}_s \mid \mathbf{X}) + \operatorname{Cov}(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad - \operatorname{Cov}(\overline{\omega}_i, \omega_{is} \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i, \overline{\omega}_i \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_i, \overline{\omega}_s \mid \mathbf{X}) - \operatorname{Cov}(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad - \operatorname{Cov}(\overline{\omega}_t, \omega_{is} \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t, \overline{\omega}_i \mid \mathbf{X}) + \operatorname{Cov}(\overline{\omega}_t, \overline{\omega}_s \mid \mathbf{X}) - \operatorname{Cov}(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}) \\
&\qquad + \operatorname{Cov}(\overline{\overline{\omega}}, \omega_{is} \mid \mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\omega}_i \mid \mathbf{X}) - \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\omega}_s \mid \mathbf{X}) + \operatorname{Cov}(\overline{\overline{\omega}}, \overline{\overline{\omega}} \mid \mathbf{X}) \Big]
\end{aligned}$$

We again calculate each term:

$$\frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}(\omega_{is}, \omega_{it} \mid \mathbf{X}) = \psi^X$$

$$\frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) = \frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}\left(\omega_{it}, \frac{1}{T} \sum_{p=-m+1}^{r} \omega_{ip} \mid \mathbf{X} \right)$$

$$= \frac{1}{T}\sigma_\omega^2 + \frac{m-1}{T}\psi^B + \frac{r}{T}\psi^X$$

$$\frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}(\omega_{it}, \overline{\omega}_s \mid \mathbf{X}) = \frac{1}{Imr} \sum_{i=1}^{I} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \operatorname{Cov}\left(\omega_{it}, \frac{1}{I} \sum_{j=1}^{I} \omega_{js} \mid \mathbf{X} \right)$$

$$
\begin{aligned}
&= \frac{1}{I}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\omega_{it}, \overline{\overline{\omega}} \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\omega_{it}, \frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{m-1}{IT}\psi^B + \frac{r}{IT}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i, \omega_{is} \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}, \omega_{is} \mid \mathbf{X}\right) \\
&= \frac{1}{T}\sigma_\omega^2 + \frac{r-1}{T}\psi^A + \frac{m}{T}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i, \overline{\omega}_i \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}, \frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq} \mid \mathbf{X}\right) \\
&= \frac{1}{T}\sigma_\omega^2 + \frac{m(m-1)}{T^2}\psi^B + \frac{r(r-1)}{T^2}\psi^A + \frac{2mr}{T^2}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i, \overline{\omega}_s \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}, \frac{1}{I}\sum_{j=1}^{I}\omega_{js} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{r-1}{IT}\psi^A + \frac{m}{IT}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip}, \frac{1}{IT}\sum_{j=1}^{I}\sum_{q=-m+1}^{r}\omega_{jq} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_t, \omega_{is} \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}, \omega_{is} \mid \mathbf{X}\right) \\
&= \frac{1}{I}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_t, \overline{\omega}_i \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}, \frac{1}{T}\sum_{p=-m+1}^{r}\omega_{ip} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{m-1}{IT}\psi^B + \frac{r}{IT}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_t, \overline{\omega}_s \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}, \frac{1}{I}\sum_{k=1}^{I}\omega_{ks} \mid \mathbf{X}\right) \\
&= \frac{1}{I}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\omega}_t, \overline{\overline{\omega}} \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{I}\sum_{j=1}^{I}\omega_{jt}, \frac{1}{IT}\sum_{k=1}^{I}\sum_{p=-m+1}\omega_{kp} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{m-1}{IT}\psi^B + \frac{r}{IT}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\overline{\omega}}, \omega_{is} \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}, \omega_{is} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{r-1}{IT}\psi^A + \frac{m}{IT}\psi^X \\
\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}(\overline{\overline{\omega}}, \overline{\omega}_i \mid \mathbf{X}) &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\mathrm{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}, \frac{1}{T}\sum_{q=-m+1}^{r}\omega_{iq} \mid \mathbf{X}\right) \\
&= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X
\end{aligned}
$$

$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\overline{\omega}},\overline{\omega}_s \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}, \frac{1}{I}\sum_{k=1}^{I}\omega_{ks} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{r-1}{IT}\psi^A + \frac{m}{IT}\psi^X$$
$$\frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\overline{\overline{\omega}},\overline{\overline{\omega}} \mid \mathbf{X}) = \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}\left(\frac{1}{IT}\sum_{j=1}^{I}\sum_{p=-m+1}^{r}\omega_{jp}, \frac{1}{IT}\sum_{k=1}^{I}\sum_{q=-m+1}^{r}\omega_{kq} \mid \mathbf{X}\right)$$
$$= \frac{1}{IT}\sigma_\omega^2 + \frac{m(m-1)}{IT^2}\psi^B + \frac{r(r-1)}{IT^2}\psi^A + \frac{2mr}{IT^2}\psi^X$$

Combining these terms and simplifying yields

$$\psi_{\hat{\omega}}^X = -\left(\frac{I-1}{IT}\right)\sigma_\omega^2 - \left(\frac{(I-1)r(m-1)}{IT^2}\right)\psi^B - \left(\frac{(I-1)m(r-1)}{IT^2}\right)\psi^A + \left(\frac{2(I-1)mr}{IT^2}\right)\psi^X \tag{B.90}$$

To summarize, Equations (B.87), (B.88), (B.89), and (B.90) express the residual-based parameters as functions of the true parameters that define the error structure. Rearranging these four equations:

$$\sigma_{\hat{\omega}}^2 = \left(\frac{I-1}{IT^2}\right)\left(T(T-1)\sigma_\omega^2 - m(m-1)\psi^B - r(r-1)\psi^A - 2mr\psi^X\right)$$
$$\psi_{\hat{\omega}}^B = \left(\frac{I-1}{IT^2}\right)\left(-T\sigma_\omega^2 + (r^2+2r+m)\psi^B + r(r-1)\psi^A - 2r^2\psi^X\right)$$
$$\psi_{\hat{\omega}}^A = \left(\frac{I-1}{IT^2}\right)\left(-T\sigma_\omega^2 + m(m-1)\psi^B + (m^2+2m+r)\psi^A - 2m^2\psi^X\right)$$
$$\psi_{\hat{\omega}}^X = \left(\frac{I-1}{IT^2}\right)\left(-T\sigma_\omega^2 - r(m-1)\psi^B - m(r-1)\psi^A + 2mr\psi^X\right)$$

In matrix notation, these four equations become:

$$\begin{bmatrix}\sigma_{\hat{\omega}}^2\\ \psi_{\hat{\omega}}^B\\ \psi_{\hat{\omega}}^A\\ \psi_{\hat{\omega}}^X\end{bmatrix} = \mathbf{\Gamma}\begin{bmatrix}\sigma_\omega^2\\ \psi^B\\ \psi^A\\ \psi^X\end{bmatrix} \tag{B.91}$$

where

$$\mathbf{\Gamma} = \frac{I-1}{IT^2}\begin{bmatrix}T(T-1) & -m(m-1) & -r(r-1) & -2mr\\ -T & r^2+2r+m & r(r-1) & -2r^2\\ -T & m(m-1) & m^2+2m+r & -2m^2\\ -T & -r(m-1) & -m(r-1) & 2mr\end{bmatrix}$$

Minimum detectable effect We are ultimately interested in deriving an expression for the MDE of an experiment as a function of the residual-based parameters, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, rather than the true parameters, σ^2_ω, ψ^B, ψ^A, and ψ^X. Recall that:

$$MDE = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}$$

Having solved for $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ as linear functions of the true parameters σ^2_ω, ψ^B, ψ^A, and ψ^X, we can define k_σ, k_B, k_A, and k_X as coefficients on the residual-based parameters $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. These coefficients will allow us to use residual-based parameters in the SCR formula in place of the true parameters. In other words, k_σ, k_B, k_A, and k_X must satisfy the following equation:[36]

$$\begin{aligned}&\left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)k_\sigma\sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right)k_B\psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right)k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}}\right]}\\ &= \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right)\sqrt{\left(\frac{1}{P(1-P)J}\right)\left[\left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\right]}\end{aligned}$$

or more simply:

$$\begin{aligned}&\left(\frac{m+r}{mr}\right)k_\sigma\sigma^2_{\hat{\omega}} + \left(\frac{m-1}{m}\right)k_B\psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right)k_A\psi^A_{\hat{\omega}} - 2k_X\psi^X_{\hat{\omega}}\\ &= \left(\frac{m+r}{mr}\right)\sigma^2_\omega + \left(\frac{m-1}{m}\right)\psi^B + \left(\frac{r-1}{r}\right)\psi^A - 2\psi^X\end{aligned}$$

We can express this equality in matrix form as:[37]

$$\begin{bmatrix}\frac{m+r}{mr} & \frac{m-1}{m} & \frac{r-1}{r} & -2\end{bmatrix}\begin{bmatrix}\sigma^2_\omega\\ \psi^B\\ \psi^A\\ \psi^X\end{bmatrix} = \begin{bmatrix}\left(\frac{m+r}{mr}\right)k_\sigma & \left(\frac{m-1}{m}\right)k_B & \left(\frac{r-1}{r}\right)k_A & -2k_X\end{bmatrix}\begin{bmatrix}\sigma^2_{\hat{\omega}}\\ \psi^B_{\hat{\omega}}\\ \psi^A_{\hat{\omega}}\\ \psi^X_{\hat{\omega}}\end{bmatrix}$$

Substituting Equation (B.91):

$$\text{(B.92)} \quad \Rightarrow \quad \begin{bmatrix}\frac{m+r}{mr} & \frac{m-1}{m} & \frac{r-1}{r} & -2\end{bmatrix}\begin{bmatrix}\sigma^2_\omega\\ \psi^B\\ \psi^A\\ \psi^X\end{bmatrix} = \begin{bmatrix}\left(\frac{m+r}{mr}\right)k_\sigma & \left(\frac{m-1}{m}\right)k_B & \left(\frac{r-1}{r}\right)k_A & -2k_X\end{bmatrix}\mathbf{\Gamma}\begin{bmatrix}\sigma^2_\omega\\ \psi^B\\ \psi^A\\ \psi^X\end{bmatrix}$$

where $\mathbf{\Gamma}$ is defined as above:

$$\mathbf{\Gamma} = \frac{I-1}{IT^2}\begin{bmatrix}T(T-1) & -m(m-1) & -r(r-1) & -2mr\\ -T & r^2+2r+m & r(r-1) & -2r^2\\ -T & m(m-1) & m^2+2m+r & -2m^2\\ -T & -r(m-1) & -m(r-1) & 2mr\end{bmatrix}$$

[36]This assumes k_σ, k_B, k_A, and k_X are functions of m, r, and I only, and do not themselves depend on σ^2_ω, ψ^B, ψ^A, or ψ^X. We show this to be true below.

[37]Note that if $m = 1$ (or $r = 1$), the corresponding ψ^B and $\psi^B_{\hat{\omega}}$ (or ψ^A and $\psi^A_{\hat{\omega}}$) parameters are undefined and no longer enter the system. Similarly, the corresponding row(s) and column(s) are removed from $\mathbf{\Gamma}$.

$\mathbf{\Gamma}$ is a singular matrix and cannot be inverted. However, we can show that, rather than having no solutions, Equation (B.92) is instead overdetermined and there are infinite solutions. We are simply interested in one such solution. To find one set of k_σ, k_B, k_A, and k_X coefficients that solve this system, we iteratively solve for each k coefficient as a function of the remaining k coefficients and then substitute into the subsequent equations. That is, we use the first equation of this system to solve for k_σ as a function of k_B, k_A, and k_X and substitute this expression in place of k_σ in the subsequent equations in the system, and we repeat for the remaining coefficients and equations. This iterative procedure yields:

$$
\begin{aligned}
k_\sigma &= \frac{I(m+r)^2 + (I-1)[r(m-1)k_B + m(r-1)k_A - 2mrk_X]}{(I-1)(m+r)(m+r-1)} \\
k_B &= \frac{I(m+r)^2(m-r+mr+r^2) + (I-1)[m^2(r-1)(2-m-r)k_A + 2mr(r-m-mr-r^2)k_X]}{(I-1)r^2(3m+r+mr+r^2-2)} \\
k_A &= \frac{I(m+r)^2 - 2(I-1)mrk_X}{2(I-1)m^2} \\
k_X &= 0
\end{aligned}
$$

We iteratively substitute each coefficient into the expressions for the remaining coefficients. That is, we first substitute this value of k_X into the expressions for the other three coefficients, then substitute k_A, and so on. This yields expressions for k_σ, k_B, k_A, and k_X in terms of m, r, and I.

$$
\begin{aligned}
k_\sigma &= \frac{I(m+r)^2}{2(I-1)mr} \\
k_B &= \frac{I(m+r)^2}{2(I-1)r^2} \\
k_A &= \frac{I(m+r)^2}{2(I-1)m^2} \\
k_X &= 0
\end{aligned}
$$

We can now express the MDE of an experiment as a function of the residual-based parameters:

$$
\begin{aligned}
MDE^{est} = \left(t^J_{1-\kappa} + t^J_{\alpha/2}\right) \Bigg\{ \left(\frac{1}{P(1-P)J}\right) \Bigg[\left(\frac{m+r}{mr}\right) \left(\frac{I(m+r)^2}{2(I-1)mr}\right) \sigma^2_{\hat{\omega}} \\
+ \left(\frac{m-1}{m}\right) \left(\frac{I(m+r)^2}{2(I-1)r^2}\right) \psi^B_{\hat{\omega}} + \left(\frac{r-1}{r}\right) \left(\frac{I(m+r)^2}{2(I-1)m^2}\right) \psi^A_{\hat{\omega}} \Bigg] \Bigg\}^{1/2}
\end{aligned}
$$

B.5.3 Estimation with ANCOVA

Here, we extend Appendices B.5.1 and B.5.2, which discuss estimating the MDE from residual-based parameters, for the ANCOVA model. As above, we do not observe the true parameters characterizing the error structure (σ^2_υ, σ^2_ω, ψ^B, ψ^A, and ψ^X). We cannot calculate these from a dataset, so we turn to residual-based versions of these parameters ($\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$), which we can use to define $MDE^{est}_{ANCOVA}(\sigma^2_{\hat{\upsilon}}, \sigma^2_{\hat{\omega}}, \psi^B_{\hat{\omega}}, \psi^A_{\hat{\omega}}, \psi^X_{\hat{\omega}}) =$

$MDE_{ANCOVA}(\sigma_\upsilon^2, \sigma_\omega^2, \psi^B, \psi^A, \psi^X)$ such that the MDE from estimated parameters is equivalent to the SCR formula's MDE derived from true parameters.

Model As in Appendix B.5.2, we assume this model is identical to the model that generates the relevant power calculation formula; in this case, SCR ANCOVA given in Equation (3.11).

That is, there are J units, P proportion of which are randomized into treatment. The researcher again collects outcome data Y_{it} for each unit i, across m pre-treatment time periods and r post-treatment time periods. For treated units, $D_{it} = 0$ in pre-treatment periods and $D_{it} = 1$ in post-treatment periods; for control units, $D_{it} = 0$ in all periods.

Assumption (Data generating process)**.** *The data are generated according to the following model:*

$$Y_{it} = \beta + \tau D_{it} + \upsilon_i + \omega_{it}$$

where Y_{it} is the outcome of interest for unit i at time t; β is the expected outcome of non-treated observations; τ is the treatment effect that is homogeneous across all units and all time periods; D_{it} is a time-varying treatment indicator; υ_i is a time-invariant unit effect distributed i.i.d. $\mathcal{N}(0, \sigma_\upsilon^2)$; and ω_{it} is an idiosyncratic error term distributed (not necessarily i.i.d.) $\mathcal{N}(0, \sigma_\omega^2)$.

Assumption (Strict exogeneity)**.** $\mathbb{E}[\omega_{it} \mid \mathbf{X}] = 0$*, where* $\mathbf{X}$ *is a full rank matrix of regressors, including a constant, the treatment indicator* $\mathbf{D}$*, and* $J-1$ *unit dummies. This again follows from random assignment of* D_{it}*.*

Assumption (Balanced panel)**.** *The number of pre-treatment observations, m, and post-treatment observations, r, is the same for each unit, and all units are observed in every time period.*

Assumption (Independence across units)**.** $\mathbb{E}[\omega_{it}\omega_{js} \mid \mathbf{X}] = 0,\ \forall\ i \neq j,\ \forall\ t, s.$

Assumption (Uniform covariance structures)**.** *Define:*

$$\psi_i^B \equiv \frac{2}{m(m-1)} \sum_{t=-m+1}^{-1} \sum_{s=t+1}^{0} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi_i^A \equiv \frac{2}{r(r-1)} \sum_{t=1}^{r-1} \sum_{s=t+1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

$$\psi_i^X \equiv \frac{1}{mr} \sum_{t=-m+1}^{0} \sum_{s=1}^{r} \text{Cov}\left(\omega_{it}, \omega_{is} \mid \mathbf{X}\right)$$

to be the average pre-treatment, post-treatment, and across-period covariance between different error terms of unit i, respectively. Using these definitions, assume that $\psi^B = \psi_i^B$, $\psi^A = \psi_i^A$, and $\psi^X = \psi_i^X$ $\forall i$.

Parameters We begin by estimating the fixed effects and residuals. Note that even though we are ultimately interested in performing a power calculation for the ANCOVA estimator, we generate residuals for this data generating process by regressing Y_{it} on a constant and unit fixed effects:

$$Y_{it} = \beta + \upsilon_i + \omega_{it}$$

For a balanced panel, the estimated coefficients are:

$$\hat{\beta} = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it}$$
$$\hat{\upsilon}_i = \frac{1}{T} \sum_{t=-m+1}^{r} Y_{it} - \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} Y_{it}$$

Then the estimated fixed effects and the residuals are

$$\hat{\upsilon}_i = (\upsilon_i - \overline{\upsilon}) + (\overline{\omega}_i - \overline{\overline{\omega}})$$
$$\hat{\omega}_{it} = \omega_{it} - \overline{\omega}_i$$

where

$$\overline{\upsilon} = \frac{1}{I} \sum_{i=1}^{I} \upsilon_i$$
$$\overline{\omega}_i = \frac{1}{T} \sum_{t=-m+1}^{r} \omega_{it}$$
$$\overline{\overline{\omega}} = \frac{1}{IT} \sum_{i=1}^{I} \sum_{t=-m+1}^{r} \omega_{it}$$

We first use the estimated fixed effects to derive an expression for $\sigma^2_{\hat{\upsilon}}$:[38]

$$\begin{aligned}
\sigma^2_{\hat{\upsilon}} &= \frac{1}{I} \sum_{i=1}^{I} \operatorname{Var}(\hat{\upsilon}_i \mid \mathbf{X}) \\
&= \frac{1}{I} \sum_{i=1}^{I} \big[\operatorname{Var}(\upsilon_i \mid \mathbf{X}) + \operatorname{Var}(\overline{\upsilon} \mid \mathbf{X}) - 2\operatorname{Cov}(\upsilon_i, \overline{\upsilon} \mid \mathbf{X}) \\
&\qquad + \operatorname{Var}(\overline{\omega}_i \mid \mathbf{X}) + \operatorname{Var}(\overline{\overline{\omega}} \mid \mathbf{X}) - 2\operatorname{Cov}(\overline{\omega}_i, \overline{\overline{\omega}} \mid \mathbf{X})\big] \\
&= \left(\frac{I-1}{IT^2}\right) \left[T^2\sigma^2_{\upsilon} + T\sigma^2_{\omega} + (m(m-1))\psi^B + (r(r-1))\psi^A + 2mr\psi^X\right]
\end{aligned}$$

[38]Note that, as stated in the strict exogeneity assumption, $\mathbf{X}$ is a full rank matrix of regressors, including a constant, the treatment indicator $\mathbf{D}$, and $J-1$ unit dummies. That is, $\mathbf{X}$ includes the covariates of the data generating process and residualizing regression, not of the ANCOVA model.

We also use the definition of residuals to derive expressions for $\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$:

$$\begin{aligned}
\sigma^2_{\hat{\omega}} &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\text{Var}(\hat{\omega}_{it} \mid \mathbf{X}) \\
&= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}[\text{Var}(\omega_{it} \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X}) - 2\,\text{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X})] \\
&= \left(\frac{T-1}{T}\right)\sigma^2_{\omega} - \left(\frac{m(m-1)}{T^2}\right)\psi^B - \left(\frac{r(r-1)}{T^2}\right)\psi^A - \left(\frac{2mr}{T^2}\right)\psi^X \\
\psi^A_{\hat{\omega}} &= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}\text{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{2}{r(r-1)I}\sum_{i=1}^{I}\sum_{t=1}^{r-1}\sum_{s=t+1}^{r}[\text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \text{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) \\
&\qquad\qquad - \text{Cov}(\omega_{is}, \overline{\omega}_i \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X})] \\
&= \frac{-1}{T}\sigma^2_{\omega} + \left(\frac{m(m-1)}{T^2}\right)\psi^B + \left(\frac{m^2+2m+r}{T^2}\right)\psi^A - \left(\frac{2m^2}{T^2}\right)\psi^X \\
\psi^B_{\hat{\omega}} &= \frac{-1}{T}\sigma^2_{\omega} + \left(\frac{r^2+2r+m}{T^2}\right)\psi^B + \left(\frac{r(r-1)}{T^2}\right)\psi^A - \left(\frac{2r^2}{T^2}\right)\psi^X \\
\psi^X_{\hat{\omega}} &= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}\text{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\
&= \frac{1}{Imr}\sum_{i=1}^{I}\sum_{t=-m+1}^{0}\sum_{s=1}^{r}[\text{Cov}(\omega_{it}, \omega_{is} \mid \mathbf{X}) - \text{Cov}(\omega_{it}, \overline{\omega}_i \mid \mathbf{X}) \\
&\qquad\qquad - \text{Cov}(\omega_{is}, \overline{\omega}_i \mid \mathbf{X}) + \text{Var}(\overline{\omega}_i \mid \mathbf{X})] \\
&= \frac{-1}{T}\sigma^2_{\omega} - \left(\frac{r(m-1)}{T^2}\right)\psi^B - \left(\frac{m(r-1)}{T^2}\right)\psi^A + \left(\frac{2mr}{T^2}\right)\psi^X
\end{aligned}$$

Minimum detectable effect Having solved for $\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$ as linear functions of the true parameters σ^2_{υ}, σ^2_{ω}, ψ^B, ψ^A, and ψ^X, we can define k_{υ}, k_{ω}, k_B, k_A, and k_X as coefficients on the residual-based parameters $\sigma^2_{\hat{\upsilon}}$, $\sigma^2_{\hat{\omega}}$, $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$. These coefficients will allow us to use residual-based parameters in the SCR formula in place of the true parameters. In other words, k_{υ}, k_{ω}, k_B, k_A, and k_X must satisfy the following equation:

$$\begin{aligned}
&(1-\theta)^2\sigma^2_{\upsilon} + \left(\frac{\theta^2}{m} + \frac{1}{r}\right)\sigma^2_{\omega} + \frac{\theta^2(m-1)}{m}\psi^B + \frac{r-1}{r}\psi^A - 2\theta\psi^X \\
&\quad = k_{\upsilon}\sigma^2_{\hat{\upsilon}} + k_{\omega}\sigma^2_{\hat{\omega}} + k_B\psi^B_{\hat{\omega}} + k_A\psi^A_{\hat{\omega}} + k_X\psi^X_{\hat{\omega}}
\end{aligned}$$

Following the solution procedure described in Appendix B.5.2, we solve for a set of k coefficients that satisfy this equality:

$$k_{\upsilon} = \left(\frac{I}{I-1}\right)(1-\theta)^2$$

$$
\begin{aligned}
k_\omega &= \left(\frac{m+\theta r}{2m^2r^2}\right)((m+r)(m+\theta r)+(1-\theta)(mr^2-m^2r)) \\
k_B &= \left(\frac{m+\theta r}{2mr^2}\right)(m-1)(m+\theta r-(1-\theta)mr) \\
k_A &= \left(\frac{m+\theta r}{2m^2r}\right)(r-1)(m+\theta r+(1-\theta)mr) \\
k_X &= 0
\end{aligned}
$$

Note, however, that these coefficients are functions of not only m, r, and I, as in Appendix B.5.2. These coefficients are also a function of θ, which is itself a function of true parameters:

$$\theta = \frac{m\sigma_\upsilon^2 + m\psi^X}{m\sigma_\upsilon^2 + \sigma_\omega^2 + (m-1)\psi^B}$$

As a result, in order to use these coefficients in a residual-based SCR ANCOVA formula, we must also derive a residual-based expression for θ. We define a set of coefficients to correspond to the numerator of θ: k_υ^N, k_ω^N, k_B^N, k_A^N, and k_X^N; we also define a set of coefficients to correspond to the denominator of θ: k_υ^D, k_ω^D, k_B^D, k_A^D, and k_X^D. We seek to find the sets of k^N and k^D coefficients that satisfy the following equality:

$$\frac{m\sigma_\upsilon^2 + m\psi^X}{m\sigma_\upsilon^2 + \sigma_\omega^2 + (m-1)\psi^B} = \frac{k_\upsilon^N\sigma_{\hat{\upsilon}}^2 + k_\omega^N\sigma_{\hat{\omega}}^2 + k_B^N\psi_{\hat{\omega}}^B + k_A^N\psi_{\hat{\omega}}^A + k_X^N\psi_{\hat{\omega}}^X}{k_\upsilon^D\sigma_{\hat{\upsilon}}^2 + k_\omega^D\sigma_{\hat{\omega}}^2 + k_B^D\psi_{\hat{\omega}}^B + k_A^D\psi_{\hat{\omega}}^A + k_X^D\psi_{\hat{\omega}}^X}$$

We again follow the solution procedure described in Appendix B.5.2 as we separately solve for a set of k^N coefficients and k^D coefficients. One such solution for the numerator coefficients is:

$$
\begin{aligned}
k_\upsilon^N &= \left(\frac{I}{I-1}\right)m \\
k_\omega^N &= \left(\frac{-1}{4r}\right)(m(m-r+2)+r(r-m+2)) \\
k_B^N &= \left(\frac{-m}{4r}\right)(m-1)(m-r+2) \\
k_A^N &= \left(\frac{-1}{4}\right)(r-1)(r-m+2) \\
k_X^N &= 0
\end{aligned}
$$

and one such solution for the denominator coefficients is:

$$\begin{aligned}
k_{\upsilon}^{D} &= \left(\frac{I}{I-1}\right) m \\
k_{\omega}^{D} &= \left(\frac{1}{2m}\right) (m(m+1) - r(m-1)) \\
k_{B}^{D} &= \left(\frac{1}{2}\right) (m+1)(m-1) \\
k_{A}^{D} &= \left(\frac{-r}{2m}\right) (m-1)(r-1) \\
k_{X}^{D} &= 0
\end{aligned}$$

Combining these two sets of coefficients, we express θ as a function of residual-based parameters:

$$\begin{aligned}
\theta = & \frac{m\left[\left(\frac{I}{I-1}\right) 4mr\sigma_{\hat{\upsilon}}^2 - (m(m-r+2) + r(r-m+2))\sigma_{\hat{\omega}}^2\right]}{2r\left[\left(\frac{I}{I-1}\right) 2m^2\sigma_{\upsilon}^2 + (m(m+1) - r(m-1))\sigma_{\hat{\omega}}^2 + m(m-1)(m+1)\psi_{\hat{\omega}}^B - r(m-1)(r-1)\psi_{\hat{\omega}}^A\right]} \\
& + \frac{m\left[-m(m-1)(m-r+2)\psi_{\hat{\omega}}^B - r(r-1)(r-m+2)\psi_{\hat{\omega}}^A\right]}{2r\left[\left(\frac{I}{I-1}\right) 2m^2\sigma_{\upsilon}^2 + (m(m+1) - r(m-1))\sigma_{\hat{\omega}}^2 + m(m-1)(m+1)\psi_{\hat{\omega}}^B - r(m-1)(r-1)\psi_{\hat{\omega}}^A\right]}
\end{aligned}$$

We can now express the SCR ANCOVA formula using only residual-based parameters:

$$\begin{aligned}
MDE_{ANCOVA}^{est} \approx (t_{1-\kappa}^{J} + t_{\alpha/2}^{J}) \times \Bigg\{ & \left(\frac{1}{P(1-P)J}\right) \\
& \times \Bigg[\left(\frac{I}{I-1}\right) (1-\theta)^2 \sigma_{\hat{\upsilon}}^2 \\
& + \left(\frac{m+\theta r}{2m^2 r^2}\right) ((m+r)(m+\theta r) + (1-\theta)(mr^2 - m^2 r))\sigma_{\hat{\omega}}^2 \\
& + \left(\frac{m+\theta r}{2mr^2}\right) (m-1)(m+\theta r - (1-\theta)mr)\psi_{\hat{\omega}}^B \\
& + \left(\frac{m+\theta r}{2m^2 r}\right) (r-1)(m+\theta r + (1-\theta)mr)\psi_{\hat{\omega}}^A \Bigg] \Bigg\}^{1/2}
\end{aligned}$$

where θ is as given above.

Estimated parameters We finally show that we can recover unbiased estimates of the residual-based parameters, as in Appendix B.5.1. We begin with the variance of the fixed effects, $\sigma_{\hat{\upsilon}}^2$, and define:

$$\tilde{\sigma}_{\hat{\upsilon}}^2 \equiv \frac{1}{I} \sum_{i=1}^{I} \left(\hat{\upsilon}_i - \overline{\hat{\upsilon}}\right)^2$$

where $\overline{\hat{\upsilon}} = \frac{1}{I}\sum_{i=1}^{I}\hat{\upsilon} = 0$. Taking expectations of both sides:

$$\begin{aligned}\mathbb{E}\left[\tilde{\sigma}^2_{\hat{\upsilon}} \mid \mathbf{X}\right] &= \frac{1}{I}\sum_{i=1}^{I}\mathbb{E}\left[\hat{\upsilon}^2 \mid \mathbf{X}\right] \\ &= \sigma^2_{\hat{\upsilon}}\end{aligned}$$

Next, to estimate the variance of the residuals, $\sigma^2_{\hat{\omega}}$, we define:

$$\tilde{\sigma}^2_{\hat{\omega}} \equiv \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\left(\hat{\omega}_{it} - \overline{\overline{\hat{\omega}}}\right)^2$$

where $\overline{\overline{\hat{\omega}}} = \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\hat{\omega}_{it} = 0$. Taking expectations of both sides:

$$\begin{aligned}\mathbb{E}\left[\tilde{\sigma}^2_{\hat{\omega}} \mid \mathbf{X}\right] &= \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=-m+1}^{r}\mathbb{E}\left[\hat{\omega}^2_{it} \mid \mathbf{X}\right] \\ &= \sigma^2_{\hat{\omega}}\end{aligned}$$

To estimate $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, we define the $[I \times 1]$ vector of residuals for period t as $\hat{\boldsymbol{\omega}}_t$ and calculate the average covariance between any two vectors in the relevant range of time periods. For the pre-period, we define:

$$\begin{aligned}\tilde{\psi}^B_{\hat{\omega}} &\equiv \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\left(\hat{\omega}_{it} - \overline{\hat{\omega}}_t\right)\left(\hat{\omega}_{is} - \overline{\hat{\omega}}_s\right) \\ &= \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\left(\hat{\omega}_{it}\hat{\omega}_{is} - \hat{\omega}_{it}\overline{\hat{\omega}}_s - \hat{\omega}_{is}\overline{\hat{\omega}}_t + \overline{\hat{\omega}}_t\overline{\hat{\omega}}_s\right) \\ &= \frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\left(\frac{I-1}{I}\right)\hat{\omega}_{it}\hat{\omega}_{is}\end{aligned}$$

Taking expectations yields:

$$\begin{aligned}\mathbb{E}\left[\tilde{\psi}^B_{\hat{\omega}} \mid \mathbf{X}\right] &= \left(\frac{I-1}{I}\right)\frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\mathbb{E}\left[\hat{\omega}_{it}\hat{\omega}_{is} \mid \mathbf{X}\right] \\ &= \left(\frac{I-1}{I}\right)\frac{2}{Im(m-1)}\sum_{t=-m+1}^{-1}\sum_{s=t+1}^{0}\sum_{i=1}^{I}\mathrm{Cov}(\hat{\omega}_{it}, \hat{\omega}_{is} \mid \mathbf{X}) \\ &= \left(\frac{I-1}{I}\right)\psi^B_{\hat{\omega}}\end{aligned}$$

Similarly:

$$\mathbb{E}\left[\tilde{\psi}^A_{\hat{\omega}} \mid \mathbf{X}\right] = \left(\frac{I-1}{I}\right)\psi^A_{\hat{\omega}}$$
$$\mathbb{E}\left[\tilde{\psi}^X_{\hat{\omega}} \mid \mathbf{X}\right] = \left(\frac{I-1}{I}\right)\psi^X_{\hat{\omega}}$$

Hence, we can recover unbiased estimates of the parameters $\sigma^2_{\hat{v}}$ and $\sigma^2_{\hat{\omega}}$ by calculating the averages of the estimated $\tilde{\sigma}^2_{\hat{v}}$ and $\tilde{\sigma}^2_{\hat{\omega}}$, respectively. However, to recover unbiased estimates of the parameters $\psi^B_{\hat{\omega}}$, $\psi^A_{\hat{\omega}}$, and $\psi^X_{\hat{\omega}}$, we must calculate $\tilde{\psi}^B_{\hat{\omega}}$, $\tilde{\psi}^A_{\hat{\omega}}$, and $\tilde{\psi}^X_{\hat{\omega}}$, respectively, and then inflate these averages by $\frac{I}{I-1}$.