Title
Essays on Energy Economics and Agricultural R&D

Author
Wang, Shanchao

Publication Date
2022

Peer reviewed|Thesis/dissertation

Essays on Energy Economics and Agricultural R&D

By

SHANCHAO WANG

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Agricultural and Resource Economics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Aaron Smith, Chair

Kevin Novan

Julian Alston

Committee in Charge

2022

Abstract

This thesis consists of three chapters. In the first two chapters, I study rebound effects in solar adoption and an energy efficiency program (air-conditioning units (AC) upgrading program). Both solar adoption and AC upgrades reduce households' energy bills and lower their average electricity prices. Households might adjust their energy consumption behavior which results in the actual reduction in energy use lower than the anticipated reduction. In chapter 1, I use novel data which contain the detailed household-level hourly purchase, sale, and solar generation in the Sacramento area. Contrary to existing literature, results do not show significant rebound effects. Using fixed effects models, I find a statistically insignificant rebound effect of 0.96%, which translates to a 0.0096 kWh increase in solar homes' total electricity consumption when the solar generation increases by 1 kWh. This effect is also economically negligible. Results from chapter 1 enrich the current literature on solar adoption rebound effects.

In contrast to chapter 1, significant rebound effects from the AC Energy Efficiency rebate program in the Sacramento Municipal Utility District (SMUD) serving area are identified in chapter 2. Household-level daily electricity consumption data and daily temperature data in the Sacramento area are utilized in this analysis. Regression mixture models are estimated by an expectation-maximization algorithm to recover premise-level temperature response functions and AC usage behavior functions. These functions are then used to calculate direct savings, total savings, and rebound effects from AC upgrades through a difference-in-difference design. On average, the AC energy efficiency rebate program reduces energy uses

in cooling by 347.10 kWh per household in one summer. The rebound effects are estimated at 20.61%. When a household saves 1 kWh in cooling from AC upgrading, its total daily electricity consumption will increase by around 0.21 kWh. The increases in consumption are mainly caused by turning on AC units more often after AC upgrades.

In chapter 3, I and coauthors study R&D lag structure in agriculture. Quite diverse models of R&D lag structures have been used by economists studying economic growth compared with those estimating returns to investments in industrial R&D or agricultural R&D. In this paper, we and coauthors empirically compare and contrast these alternative models and their implications for R&D knowledge stocks using data on multifactor productivity (MFP) in U.S. agriculture and U.S. public agricultural R&D investments. We employ a model selection procedure based on a combination of time-series properties of data, econometric estimation performance, and consistency of estimates with strongly held economic priors. We reject the models used in studies of economic growth and industrial R&D both on prior grounds and using statistical tests. The preferred model is a 50-year gamma lag distribution model similar in shape to Huffman and Evenson's (1993) trapezoidal lag model. In this gamma lag model the effects of an investment in agricultural R&D on the R&D knowledge stock rise to a peak after 13 years and are mostly dissipated after 35 years. The estimated elasticity of MFP with respect to the knowledge stock is 0.28 and the implied marginal benefit-cost ratio is 23:1.

Key words: rebound effects, solar adoption, energy efficiency programs, agricultural productivity, R&D lags, model selection, rate of return

JEL Codes: D12, D24, E23, O31, 047, Q16, Q40, Q42

Acknowledgments

Firstly, I would like to express my sincere gratitude to my advisor Prof. Smith for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Novan and Prof. Alston, for their insightful comments and encouragement, and the opportunities of co-authorship. Special thanks go to Prof. Alston for his prompt replies and flexible meeting schedules.

My sincere thanks also go to Prof. Sexton, and Prof. Agerton, who provided me an opportunity to explore topics related to natural resources and dynamic optimization. These pieces of knowledge are precious and can be great treasures for my future work.

I thank my colleagues for the inspiring discussions, for the encouragement they gave when I met difficulties, and for all the fun we have had in the last five years. In particular, I am grateful to Yijing Wang for supporting me emotionally.

Last but not the least, I would like to thank my family: my parents and to my grandparents for supporting me spiritually throughout writing this thesis and my my life in general.

Contents

List of Figures

List of Tables

Essay 1

Does electricity consumption respond to solar adoption? New evidence from smart meter data

1.1 Introduction

With concern about climate change and fears of depleting fossil fuels, governments and private institutions have been actively investing in renewable energy. Solar power, as a promising type of renewable energy, has rapidly expanded in the United States. In recent years, due to the decreasing installation costs and significant subsidies, rooftop solar generation grew at a rate of roughly 22% per year from 2017 through 2019 (US EIA, 2020*a*). From 2016 to 2019, California added 5.3 GW of new rooftop solar photovoltaic (PV) capacity, more than any other states in the US (US EIA, 2020*b*). Experts in the US Energy Information Administration (EIA) expect California to continue to be the leading state in rooftop solar production, especially under the regulation that the majority of newly built houses must install a solar PV system. This new building code has been effective since January 1st, 2020.

Solar panel adoption is advertised as an effective way to reduce family energy bills. For a household, if the total consumption is fixed, then the electricity generated from solar

panels will simply substitute for purchases from electric utilities and hence save the expenditure. However, this calculation is likely to overstate the reduction in electricity purchase. Households substitute solar generated electricity for purchased electricity, and as purchases decrease, the average price that households pay for purchased electricity also decreases. The implications of a lower average price are twofold. First, from a consumer's perspective, the economic theory predicts that a lower price increase electricity consumption. This induced increase in consumption is documented as the "rebound effects" in the energy economics literature.[1] Second, for electric utilities, the decrease in electricity purchases induces them to raise per kWh rates to cover their costs. This adversely affects all households that purchase electricity from the electric utility, including those who do not consume solar energy, and these households are generally associated with lower socio-income status than solar consumers (Liang et al., 2018; Rohan, Paul and Shuhei, 2019).

Recent literature has found significant increases in households' electricity consumption post solar adoption. Qiu, Kahn and Xing (2019) use household level hourly solar generation and daily electricity consumption data in Phoenix Arizona to quantify the rebound effect. They conclude that an increase of 1 kWh in solar electricity generation could significantly increase the households total consumption by 0.18 kWh (the rebound effects are hence estimated at 18%). Two earlier studies (Deng and Newton, 2017; Havas et al., 2015) attempt to estimate rebound effects utilizing household level data in Australia. The estimated effects in these two studies are 20% and 15% respectively. More recent work done by Beppler, Matisoff and Oliver (2021) also uses household-level data and finds a 28% rebound effect. This is the largest rebound effect that has been found so far in the solar energy literature.

Various mechanisms may explain the significantly positive rebound effects post PV unit adoption apart from the price effects that we mentioned above. For example, households may act exceptionally rationally after PV installation so they are moving to the lower price

[1]In the energy economics, rebound effects refer to the situation in which the marginal cost of an energy service decreases, and the consumption of that service increases. Since we are not directly examining the consumption in solar energy uses, the rebound effects in this study are actually indirect.

tier. With a lower marginal cost of consuming energy services, they increase their electricity consumption; households bump up their electricity consumption (e.g., turning on AC units more often) but do not see dramatic increases in their monthly bills due to solar generation. Therefore, they increase electricity consumption even more; households might consume more electricity post PV unit adoption if they purchase or co-adopt other devices (e.g., electric vehicles) will result in increases in the households' electricity consumption. However, these hypotheses are generally untestable due to data limitation (e.g., we do not have data on co-adoption for solar adopters). Moreover, price and income effects alone are not convincing enough to explain the magnitude of rebound effects found in previous studies given how insensitive households electricity consumption has typically been found to be in response to price and income (Zhu et al., 2018). Therefore, we cast doubt on and reexamine the large rebound effects documented in the current literature.

Our study focuses on understanding the size of rebound effects after the adoption of solar panels using new household-level data in the Sacramento Municipal electricity District (SMUD). In contrast to previous results, we find the rebound effect is neither economically nor statistically significant. Several distinctive features of the new data make our results more convincing than those in previous studies.

Compared to Deng and Newton (2017) and Havas et al. (2015), in which the authors study adoption of solar panels in Australia, we focus on the Sacramento energy market, and our results may be easier to generalize to other areas in the US. There are two main types of solar feed-in tariffs, namely gross feed-in tariff and net feed-in tariff (also known as Net Energy Metering (NEM) or Net Metering (NM)). Some areas in Australia (including areas studied by Deng and Newton (2017) and Havas et al. (2015)) implement a gross feed-in tariff where households sell all the solar energy they generate to the grid and purchase all the electricity they consume from utilities. In the US, where a net feed-in tariff is applied, households can purchase, generate and sell electricity all at the same time. The difference between the two tariffs — comparing the United States and Australia — can give rise to

different behaviors in energy use (Beppler, Matisoff and Oliver, 2021). The dominant cause of rebound for the gross feed-in tariff is households' perceived increase in "income" from selling generated electricity at a flat rate while for net feed-in tariff users, the main cause is that households consume electricity at a lower perceived marginal price. Although theory suggests consumers react to marginal prices, empirical studies find that customers tend to respond to the lump sum bills instead of current marginal prices (Ito, 2014). Therefore, we might see lagged or smaller rebound effects when studying NM systems.

Qiu, Kahn and Xing (2019), selected a set of solar adopters (277 households) to compare with around 4000 non-solar households, while our data contain early adopters, adopters, and non-adopters in the SMUD area (more than 12,000 households) during 2017 and 2018. Moreover, in their main specification, Qiu, Kahn and Xing (2019) regress daily consumption on daily solar generation and other fixed effects, which reflects marginal daily consumption changes with respect to changes in solar generation resulting from changes in solar irradiation. In our analysis, we explore the changes at the margin (in the robustness check section) as well as total changes in daily consumption pre- and post- solar panel adoption. In this sense, our study is more comprehensive.

Finally, Beppler, Matisoff and Oliver (2021) imputed solar generation data from each household's installed solar system size instead of directly observing it. If solar generation is overestimated, the rebound effect will be overstated as increases in daily consumption are inflated by high solar generation estimates. For example, suppose a household consumes 25 kWh per day before solar panel adoption. After installation, the household purchases 20 kWh from the utility, and the solar panel generates 5 kWh, all of which is consumed. Then the real total change in daily consumption is 0 kWh. Now if solar generation is overestimated as 7 kWh per day. The total consumption post-adoption is estimated at 27 kWh, so the estimated daily consumption goes up by 2 kWh. Dividing this estimated increase by the estimated generation of 7 kWh, we get an estimated rebound effect of around 28.57%. Our analysis avoids this potential pitfall by directly observing hourly purchase, sale, and

generation, hence providing more accurate results.

In what follows, we first give a detailed description of our data including the data cleaning process and some graphical evidence. In section 1.3, we present our identification strategies followed by results from our novel data. We also perform several robustness checks in section 1.5. We conclude in section 1.6.

1.2 Data

This paper combines household-level hourly electricity purchasing, selling, and solar production data in the Sacramento Municipal electricity District (SMUD) service area for the period January 1, 2017 – December 31, 2018. We also utilize daily average weather data (e.g. dry-bulb temperature and visibility) from a National Oceanic and Atmospheric Administration (NOAA) station at the Sacramento International Airport to examine whether households responded to outdoor temperature differently before and after the installation of solar panels.

The SMUD electricity data encompasses both commercial and residential premises. In Figure 1.1, we plot the average hourly purchase, sale, and generation across premises by different premises types. Although commercial and residential premises have similar hourly average distributions in solar generation and sale, the distribution of electricity purchases by commercial premises across a day is much smoother than that of non-commercial households. Moreover, compared with residential premises, commercial premises use much more electricity. To avoid confounding the consumption differences before and after solar panel installation with the consumption differences between commercial and residential premises, we focus exclusively on non-commercial single-family households.

Households without solar panels were randomly chosen from the SMUD service area. For households with solar panels, the following procedure was applied: at the end of each year (December 31 in 2017 and 2018), households with solar panels installed were selected and their daily usage data in that year were extracted. For example, if a household adopted PV units in June 2017, then we will have its daily electricity data in 2017 and 2018. However,

if a household install solar panels in June 2018, we do not observe its data for 2017. We will only observe its data for 2018. Therefore, we decided to focus on premises that installed solar panels in 2017 and have full-year data in 2018 so that the same days in a year pre- and post-adoption are included.

In Figure 1.2, we plot the average daily usage across households (defined as electricity purchase from utility plus solar generation minus generated electricity sold back to the grid) in weeks pre- and post-installation. The grey bar indicates 2.5% and 97.5% percentiles of average daily usage in a week across all households. We defined the first date when a household's solar generation turns from 0 to a positive number as the installation date. The installation date and the trailing 6 days are defined as week 1 post-installation, the next 7 days are defined as week 2 post-installation, and so on. Similarly, 7 days prior to the installation date are defined as week -1 (pre-installation week 1), and so on.

We see a clear difference in terms of the average daily usage and quantiles pre- and post-adoption. The post-installation daily average usage is continuous, and the quantiles are fairly similar across weeks, while the pre-adoption average daily usage is scattered and shows a high variation in quantiles.

One caveat of our data is that the SMUD provided the premises-specific data on electricity purchase, sale, and generation in three separate datasets. When we combine these three datasets, purchase and sale are merged perfectly. However, there is a large number of null values in the solar generation when we combine the generation data with the merged purchase-sale data. We infer these null values as 0 since all corresponding hourly sales are also 0. After this data cleaning process, we recreated Figure 1.2 and the results are shown in Figure 1.3.[2]

As shown in Figure 1.3, we now see the means of daily average usage are continuous pre- and post-installation. Moreover, the quantiles look more reasonable than before converting null values to 0s in the solar generation. All households in our sample have full-year data, i.e.,

[2]In Figure 1.2, we do not include premises-date entries when there are null values in solar generation. The procedure described above expands our data panel by adding more entries to pre-installation dates.

data for 730 days. We further exclude households that have a pre-adoption period of fewer than 30 days and households with more than 10 days with 0 solar generation post-adoption. This ensures we have enough pre-adoption data for each premises and exclude "dropouts".

We categorize households with solar panel into two groups: adopters and early adopters. Early adopters are defined as households that installed solar panels before January 1, 2017, while adopters are defined as households that installed solar during 2017. As described earlier, we also acquire data on a random sample of households in the SMUD service area that did not have solar panels during 2017 or 2018.

Summary statistics of different groups are listed in Table 1.1. As before, we define daily consumption as daily electricity purchases from SMUD plus solar generation minus sell-backs. From panel A, it is clear that early adopters had the highest average daily consumption while non-adopters had the lowest. We also check the pre- and post- statistics for adopters. We do not see obvious changes in the mean and median.

In Figure 1.4, we plot the average hourly consumption across premises for adopters before and after solar panel installation. For a single hour, the difference is also insubstantial. Notice that so far we have not controlled for temperatures and dates of installation. In the next section, we use regressions to pin down the effects of solar installation on household consumption, with temperatures controlled.

1.3 Identification Strategy

1.3.1 Baseline model

We start with the following baseline model to estimate the effect of solar panel installation on household electricity consumption:

$$Con_{i,d} = \alpha_i + \beta Inst_{i,d} + \gamma_1 \cdot Temp_d + \gamma_2 \cdot S_d + \varepsilon_{i,d} \tag{1.1}$$

where i indexes the premises and d indexes the dates. $Con_{i,d}$ is the total daily consumption on date d. α_i is the household fixed effect that captures the time-invariant household

unobservables that may affect household electricity consumption. For example, this term can encompass the location or size of a household which may affect its electricity consumption (e.g. a house surrounded by trees may consume less energy during summer). In the robustness checks section, we add alternative fixed effects and discuss the results. $Inst_{i,d}$ is a premises-specific solar panel installation status dummy variable. It switches from 0 to 1 on the date when the premises starts to generate solar electricity. $Temp_d$ is the average daily air temperature on date d while S_d is a non-linear transformation of $Temp_d$.[3] $\varepsilon_{i,d}$ is the error term. Our main parameter of interest is β which accounts for the average consumption difference between premises with and without solar panels.

As shown in Figure 1.5, daily consumption respond non-linearly to air temperature. Following Novan, Smith and Zhou (2022), we choose a restricted cubic spline model with three knots to reflect this non-linear relationship. S_d is constructed as:

$$S_d = (Temp_d - Q_1)_+^3 - \frac{Q_3 - Q_1}{Q_3 - Q_{min}} \cdot (Temp_d - Q_{min})_+^3 + \frac{Q_{min} - Q_1}{Q_3 - Q_{min}} \cdot (Temp_d - Q_3)_+^3 \quad (1.2)$$

where Q_1 and Q_3 are the first and the third quartiles of observed air temperature in the sample. Q_{min} is the temperature corresponds to the minimum average consumption. $(x)_+$ is a function that gives x when $x > 0$ and 0 otherwise. In our sample, the Q_1, Q_{min}, Q_3 are 51, 61 and 70°F respectively. [4]

To ensure the estimation of β is unbiased, we assume that other non-temperature drivers of demand for electricity are not systematically changing around the time PV units are adopted. This is a strong assumption given that we do not observe whether there were ongoing major home improvement projects when households adopted solar. However, the assumption is plausible given that: our sample size for households that adopted solar panels during 2017 is large (1659 premises). It is unlikely that all these households would conduct

[3]We obtain hourly dry bulb temperature data from NOAA and take averages of hourly temperature to get daily averages. Daily average temperatures are rounded to the closest integers.

[4]While 61°F is the temperature when minimum average electricity consumption occurs, 51°F and 70°F are the 25% and 70% quantiles of the temperature distribution in 2017 to 2018.

major home improvement projects when they install PV units. In the case that a group of households did make home improvements, our estimate of β can be either upward or downward biased within a household. For example, a household could install other energy efficiency appliances while its installed PV units. Then our β estimate for this kind of household will be underestimated as we attribute the energy savings from appliances to solar installation. On the contrary, if a household installed energy intense appliances, we would overestimate β. It is impossible to determine the direction of bias without additional information. Hence we assume the two cases mentioned above are rare and cancel out each other when we run the baseline model on all households.

1.3.2 The role of temperature

In equation 1.1, we assume that all types of premises have the same temperature response function (i.e., households do not change their cooling/heating behaviors after solar adoption). However, as indicated by literature (Deng and Newton (2017), Qiu, Kahn and Xing (2019), and Beppler, Matisoff and Oliver (2021)), households may change their behavior because of income effects (i.e., they consider the solar generation and sell-backs as a form of income, and consequently turn on cooling and heating systems more often). To test this, we add interaction terms between solar panel installation and temperature. Specifically, we are interested in estimating the following model:

$$Con_{i,d} = \alpha_i + \beta_1 Inst_{i,d} + \gamma_1 Temp_d + \gamma_2 S_d + \beta_2 Inst_{i,d} \cdot Temp_d + \beta_3 Inst_{i,d} \cdot S_d + \varepsilon_{i,d} \quad (1.3)$$

If the installation of solar panels changes consumer behavior in using cooling and heating systems, we would expect the coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ to be significant.

1.3.3 Hourly model

One distinctive feature of our data is the observation of hourly consumption enables us to observe how changes in consumption are distributed throughout a day. To do this, we reestimate equation (1.1) by replacing the dependent variable with hourly consumption.

That is:

$$Con_{i,d}^{h} = \alpha_i^h + \beta^h Inst_{i,d} + \gamma_1^h \cdot Temp_d + \gamma_2^h \cdot S_d + \varepsilon_{i,d}^h \tag{1.4}$$

Notation is the same as in equation (1.1), except for the left-hand side variable. $Con_{i,d}^{h}$ is hourly electricity consumption for household i, on date d hour h, where $h \in [1..24]$. $\varepsilon_{i,d}^h$ is the corresponding error term. We estimate 24 versions of equation (1.4), so the coefficients (e.g. β^h) are hour specific. Notice that we use the average daily temperature instead of hourly temperature in the hourly consumption regressions. This is because hourly temperature and average daily temperature are highly correlated in the Sacramento area. Using the hourly temperature will not affect our results and insights. Our parameter of interest is still $Inst_{i,d}$ which captures the difference in household electricity consumption pre- and post-adoption.

1.4 Results

In this section, we present results from the models described in the previous section. We start with our baseline model in equation (1.1). In Table 1.2, we run the model on the full sample (column (1)) and the adopter-only sample (column (2)).

The coefficients on the installation dummies are 0.175 and 0.214 for the full sample and adopter-only sample, respectively. We can interpret these estimates as: on average, solar adoption increases a household's daily consumption by 0.175 kWh for all adopters and 0.214 kWh for early adopters who installed solar panels during 2017. Both estimates are statistically insignificant even at a 10% level of significance. Standard errors are clustered at the household level and the week-year level to adjust for autocorrelation of the same household across different weeks, and the correlation among households in the same week.

Once we obtain an estimate of the change in consumption associated with solar adoption, we can calculate rebound effects by dividing the change by median solar generation in the full sample and the adopter-only sample.[5] The rebound effects are estimated at 0.96% and 1.17% for the full sample and adopter-only sample. These results differ from previously published

[5]We use the median solar generation instead of the mean solar generation since the generation data are skewed right. The median is less sensitive to extremely high solar generation.

estimates. As described, our data are unique in the way that it includes observations of hourly purchases, sale, and generation which enables us to estimate the rebound effects precisely.

We also calculate two versions of treatment effect on treated based on different samples. The consumption change estimates are divided by the median daily consumption of all households without solar panels (non-adopters and adopters pre-adoption, denoted as TOT1) and the median daily consumption for solar adopters before solar installation (denoted as TOT2). The TOT1 and TOT2 are the same for the adopters-only sample as the denominators are the same (both are median daily consumption across households in pre-adoption days). The TOT1 and TOT2 for the full sample are 0.84% and 0.73%. This means after solar panel installation, adopters on average increase their daily consumption by 0.84%, and households who installed solar panels during 2017 increase their daily consumption by 0.73%.

In Table 1.3, we present results from running equation (1.3) on the adopter-only samples. The interaction terms and the installation dummy jointly estimate the differences in the temperature response function pre- and post- solar adoption. We do not run the equation on the full sample since the interaction terms would then also capture the differences in temperature response functions between non-solar adopters and early adopters (more on this later).

As discussed earlier, if solar-adoption changes the household's cooling and heating behavior, we would expect the coefficients on the interaction terms (i.e., Inst$\times$Temp, and Inst$\times$S) to be significant. However, neither of these two estimates is significant at the 10% significance level.

Since we used a non-linear transformation of temperature in our model, it is not straightforward to see the differences in temperature response functions under various temperatures from the estimates shown in Table 1.3. We visualize differences in the temperature response function conditional on the average daily temperature. Given equation (1.3), for each premises i on date d we have predicted consumption with solar panels,

$\widehat{Con_{i,d}}|_{Inst_{i,d}=1} = \hat{\alpha}_i + \hat{\beta}_1 + \hat{\gamma_1} Temp_d + \hat{\gamma_2} S_t + \hat{\beta}_2 Temp_t + \hat{\beta}_3 S_t$ and without solar panels, $\widehat{Con_{i,d}}|_{Inst_{i,d}=0} = \hat{\alpha}_i + \hat{\gamma_1} Temp_d + \hat{\gamma_2} S_d$. The predicted difference in consumption with versus without solar panels is then $\hat{\beta}_1 + \hat{\beta}_2 Temp_t + \hat{\beta}_3 S_t$. Notice that since we do not consider heterogeneous responses (i.e., all terms do not interacted with household fixed effects), the differences are the same across households, conditional on temperature.

Results are plotted in Figure 1.6. The solid black line plot $\hat{\beta}_1 + \hat{\beta}_2 Temp_d + \hat{\beta}_3 S_d$ against daily average temperature. Coefficients estimates are from fixed effects models in Table 1.3. The grey areas around the solid line are 1.96 times the standard deviation from the solid lines.[6]

The differences in daily consumption are slightly negative for temperatures above 60°F, suggesting that solar adoption reduces a household's cooling behavior on average. For temperatures below 60°F, the differences are positive. However, differences under various temperatures are all insignificant at the 5% significance level. We would interpret results below 45°F with caution since there were only 19 and 15 days in 2017 and 2018 when the daily average temperature fell below 45°F.[7]

To further confirm that the temperature response functions are not affected by solar adoption, we run the following regression for each household in our sample:

$$Con_{i,d} = \gamma_{1,i} \cdot Temp_d + \gamma_{2,i} \cdot S_d + \varepsilon_{i,d} \tag{1.5}$$

This is the household-level temperature response function. For adopters, we divide their data into two parts: dates before and dates after the solar panel installation, and run equation (1.5). Denote the two parts of the data as adopter-pre and adopter-post. After running regressions for each household, point estimates $\hat{\gamma_{1,i}}$, and $\hat{\gamma_{2,i}}$ are averaged by household

[6]The standard deviation is calculated as the square root of $var(\hat{\beta}_1 + \hat{\beta}_2 Temp_d + \hat{\beta}_3 S_d) = var(\hat{\beta}_1) + Temp_d^2 \cdot var(\hat{\beta}_2) + S_d^2 \cdot var(\hat{\beta}_3) + 2 \cdot Temp_d \cdot cov(\beta_1, \beta_2) + 2 \cdot S_d \cdot cov(\beta_1, \beta_3) + 2 \cdot Temp_d \cdot S_d \cdot cov(\beta_2, \beta_3)$. The variance-covariance matrices are adjusted by considering correlations within households and week by year.

[7]The same is true for extremely high temperatures such as above 85°F. There were 23 and 2 such days in 2017 and 2018, respectively.

types (i.e., adopters, non-adopters, adopter-pre, and adopter-post) to provide the average temperature response function.

Results are plotted in Figure 1.7. The colored areas are 95% confidence intervals of temperature response functions. The first thing to notice is that early adopters have the highest predicted daily consumption regardless of temperatures. This aligns with the summary statistics in Table 1.1. It is also the reason why when we run the model with installation-temperature interaction terms (equation (1.3)), we exclude early adopters and non-adopters to avoid confounding effects of solar installation with differences in temperature response functions across household types. Comparing the adopter-pre and adopter-post temperature response functions, the two functions diverge after 65°F, which contradicts what we find in Figure 1.6, where we observe no differences between pre- and post-adoption periods.

The reason behind seemingly contradictory results is the structure of our data. As described in the data section, we observe adopter households that installed solar panels during 2017 and keep tracking their daily utility profiles in 2018. Adopters installed solar panels on different dates in 2017, and some adopters have shorter pre-adoption periods than others. If a household installed solar panels in April 2017, it is very likely the temperature response function will be downward sloping and nearly flat when the temperature is beyond 65°F since the temperatures in the first four months of 2017 were all below 65°F. To get predicted daily consumption, the model will make extrapolation and hence flatten the curve beyond 65°F. We plot the temperature distribution of our data by household groups in Figure 1.8. Indeed, the adopter pre sample has higher density in lower temperatures while the other groups share a similar distribution. This fact is not considered in Figure 1.7.

To take into account the effects of installation dates, we restrict adopters to those who adopted solar on or after July 1, 2017. The cutoff date is chosen since the pre-adoption period (January 1, 2017 to June 30, 2017) of these adopters has a full set of temperatures ranging from 39°F to 90°F (see Figure 1.9 in which we plot the daily average temperature in 2017).

In Figure 1.10, we recreate temperature response functions as in Figure 1.7 but exclude adopters who installed solar panels before July 1, 2017. Now the two functions of adopter-pre and adopter-post almost overlap each other. At average daily temperatures below 60°F, adopter-post has higher daily consumption than adopter-pre, but when temperature increases, adopter-post has slightly lower consumption than adopter-pre, although these differences are insignificant at a 5% significance level. This is the same result that we obtained earlier from Figure 1.6. We confirm that there is no change in the household's cooling and heating behavior after solar adoption.

We also check the rebound effects of solar adoption on hourly consumption across all hours of the day. After running 24 versions of equation (1.4) on the full sample, we saved point estimates of β^h's and the corresponding 95% confidence intervals. Results are visualized in Figure 1.11. The changes in consumption are also not statistically significant at any hour, although we see the point estimate reaches its peak at 6 pm.

1.5 Robustness Checks

In this section, we alter some of the specifications in our settings with alternative methods/-parameterizations. This section serves as a complement to our main results and checks their robustness.

1.5.1 Fixed effects

In our main specification, we allow for individual fixed effects which capture the average effects of time-invariant unobserved variables such as household characteristics (e.g. areas). Here we include additional fixed effects in our baseline model. The year fixed effects absorb any annual common shock that is universal to all households in the SMUD service area. When year fixed effects interact with premises fixed effects, the interaction terms will capture household-year specific common trends, and these trends are different across premises. For example, this could be the number of days of home occupancy that we do not observe. Finally, we include a set of time-related terms as described by Qiu, Kahn and Xing (2019), which include a federal holiday indicator (a dummy variable indicating if a date is a federal

holiday or not), day of week, day of month, and month of year fixed effects. The time-related fixed effects control for time-varying factors that are common across households. For example, promotions by SMUD or change of energy policy in California.

Results from alternative fixed effects are presented in Table 1.4 and Table 1.5 for the full sample and the adopter-only sample. For comparison, we also include the main results in column (1) of both tables. For the full sample, the largest rebound effect occurs when household and year fixed effects are included. However, the rebound effect is still statistically insignificant at the 10% significance level. Comparing the results in columns (3) and (4) of Table 1.4 and Table 1.5, the change in our main parameter of interest is inconsequential. For the adopter-only sample, we also find similar results. Across both samples with various alternative fixed effects, the coefficients on the temperature-related terms are all significant as temperature is the main factor that affects daily electricity consumption.

1.5.2 Choice of knots

Different knots are chosen to transform temperatures in the baseline model (i.e. equation (1.1)). Instead of picking the corresponding temperature for minimum average consumption, 25% and 75% percentiles, we run two additional specifications. Column (2) in Table 1.6 and Table 1.7 uses the 10%, 50% and 90% quantiles of temperature distribution in two years (i.e., 2017 and 2018), while Column (3) uses five knots located at the 10%, 30%, 50%, 70%, and 90% quantiles of the distribution. Column (1) in both tables present results from our baseline model.

From the point estimates of columns (2) and (3), regardless of whether the full or the adopter-only sample, when we include more knots, rebound effects increase. For the full sample, the rebound effect increases from 0.175 to 0.614, and from 0.214 to 0.64 for the adopter-only sample. Although we see increases in the point estimates, these rebound effects do not significantly differ from 0.

1.5.3 Linear spline transformation

We chose the restricted cubic spline transformation to capture the non-linear relationship between temperatures and the daily average electricity consumption. Novan and Smith (2018) use piece-wise linear splines to model the temperature-consumption relationship. Contrary to the restricted cubic spline, piece-wise linear splines are less flexible. However, highly flexible functions may absorb too much of the variation in the consumption and leave little room for the baseline model to capture the rebound effects.

Therefore, in this section we replace restricted cubic spline transformation with piece-wise linear transformation. In particular, we estimate the following regression on the full sample and the adopter-only sample:

$$Con_{i,d} = \alpha_i + \beta_1 \cdot Inst_{i,d} + \gamma \cdot W + \varepsilon_{i,d} \tag{1.6}$$

As in equation (1.1), $Inst_{i,d}$ is the installation dummy turns from 0 to 1 when a premise first starts to generate solar electricity. W is a column vector of transformations of temperature $Temp_d$, and γ is a row vector of parameters. Suppose we have m knots (denoted as $k_i,\ i \in \{1, ..., m\}$) for the piece-wise linear function. Then W contains $m + 1$ elements.

$$W = \begin{bmatrix} min(Temp_d, k_1) \\ ... \\ min(max(Temp_d - k_{i-1}, 0), k_i - k_{i-1}) \\ ... \\ max(Temp_d - k_m, 0) \end{bmatrix}$$

We set knots at 51, 61, and 70°F as we did for the restricted cubic spline. Three knots will introduce four linear segments in the temperature response function. Below are examples of

the first two segments. Other segments are analogues of these two.

$$\begin{aligned} Temp_d \leq 51 \quad & Con_{i,d} = \alpha_i + \beta_1 \cdot Inst_{i,d} + \gamma_{1,1} \cdot T_d \\ 51 < Temp_d \leq 61 \quad & Con_{i,d} = \alpha_i + \beta_1 \cdot Inst_{i,t} + \gamma_{1,1} \cdot 51 + \gamma_{1,2} \cdot (T_d - 51) \end{aligned}$$

Notice that we do not include temperature itself in equation (1.6). Including temperature $Temp_d$ will result in multicollinearity given the linear transformation nature of the piece-wise linear splines.

Regression results are presented in Table 1.8 for the full sample (column (1)) and the adopter-only sample (column (2)). All of the coefficients of linear spline variables are statistically significant at the 1% significance level, while the rebound effects are insignificant in both samples.

1.5.4 Overall and marginal effects

Although the time-invariant characteristics are captured by household fixed effects, one drawback of our data is that we do not observe the sizes of households' solar panels. Another limitation of our baseline model (equation (1.1)) is that we only investigate the overall rebound effects on consumption after solar adoption. Qiu, Kahn and Xing (2019) study consumption responses at the margin — that is, how daily consumption responds to marginal increases in solar generation. In this section, we incorporate marginal rebound effects into our baseline model while controlling for solar system sizes. Specifically, we run the following model on the full sample and the adopter-only sample:

$$Con_{i,d} = \alpha_i + \beta_1 Inst_{i,d} + \gamma_1 \cdot Temp_d + \gamma_2 \cdot S_d + \gamma_3 Size_{i,d} + \gamma_4 Gen_{i,d} + \varepsilon_{i,d} \tag{1.7}$$

Notation is the same as in equation (1.1) except that we include two additional variables: $Size_{i,d}$ and $Gen_{i,d}$.

Measures of $Size_{i,d}$ are constructed from observations of the average solar generation for household i in the post-adoption period. This variable is a constant for early adopters, 0 for

non-adopters, and all 0 in the pre-adoption period and a constant in the post-adoption period for adopters. In other words, we use average solar generation as a proxy for solar system size. The coefficient of this variable captures the effect of solar system size on rebound effects. $Gen_{i,d}$ created by subtracting $Size_{i,d}$ from the solar generation of household i on date d. For early adopters and adopters in the post-adoption periods, this variable is the demeaned solar generation. For non-adopters and adopters in the pre-adoption periods, this variable is all 0. The coefficient on $Gen_{i,d}$ can be interpreted as the marginal effect of rebound effects.

Table 1.9 summarizes results from estimating equation (1.7). Regardless of samples, the coefficients before the installation dummy, the solar system size proxy, and the demeaned solar generation are all insignificant at the 10% significance level. This means there are no overall or marginal rebound effects. Although the point estimates of the consumption changes in response to solar adoption (i.e., coefficients of $Inst_{i,d}$) increase to 0.687 for the full sample and 0.742 for the adopter-only sample, the clustered standard errors for these estimates also grow dramatically.

1.6 Conclusion

We investigate the impact of solar adoption on household electricity consumption in the SMUD service area. This study uses a novel dataset which contains the detailed observations of household-level purchase, sale, and generation. In contrast to previous studies, we do not find significant changes in household electricity consumption before and after the adoption of solar panels. Our findings are robust to various alternative specifications.

Contributions of our study are twofold. First, our sample not only contains a large number of early adopters, non-adopters, and adopters but also complete electricity profiles, especially in solar generation. Previous studies either rely on small and selected solar adopters or on estimated solar generation. Our data fill the gap and our analysis better reveals the households' actual responses to solar adoption. Second, while rebound effects are important when considering energy policy, we do not find sizable or statistically significant rebound effects for solar adoption.

We also realize our analysis has some limitations. First, we do not observe households' billing information. Empirical studies have shown that customers tend to respond to lump sum bills, or lagged average prices (Ito, 2014). Average prices varies across both households and time, which cannot be captured by the fixed effects that we specified in the robustness check section.[8] Second, we observe a relatively short panel compared to previous literature. Qiu, Kahn and Xing (2019) use data from 2013 to 2017, while Beppler, Matisoff and Oliver (2021) use data trimmed to December 2010 through June 2018. It might take more than one year for rebound effects to become noticeable (which we are unable to detect from our data), although Beppler, Matisoff and Oliver (2021) find significant rebound effects just one month post solar installation. However, given the size of our sample and various robustness checks that we conducted, these limitations do not undermine our conclusion that, in our case study, there were no rebound effects for solar adoption in the one to two years post-adoption.

Another interesting finding in our analysis is the differences in consumption levels for early adopters, non-adopters, and adopters (see Table 1.1 and Figure 1.7). This seems to confirm that solar adoption is associated with higher socio-income status as noted by Liang et al. (2018). Even though we do not find rebound effects, subsidies paid to solar adopters may harm lower socio-income households as utilities need to increase electricity retail prices in order to compensate for losses from reduced purchases by solar adopters.

[8]Although the marginal prices (retail prices set by the utility companies) are the same for some households in given period, solar adopters also sell their solar generation back to grids which brings down the average prices faced by households. Moreover, each household can have a different average price due to the heterogeneity in solar generation and consumption.

References

Beppler, Ross C., Daniel C. Matisoff, and Matthew E. Oliver. 2021. "Electricity consumption changes following solar adoption: Testing for a solar rebound."

Deng, Gary, and Peter Newton. 2017. "Assessing the impact of solar PV on doemestic electricity consumption: Exploring the prospect of rebound effects." *Energy Policy*, 110: 313–324.

Havas, Lisa, Julie Ballweg, Chris Penna, and Digby Race. 2015. "Power to change: Analysis of household participation in a renewable energy and energy efficiency programme in Central Australia." *Energy Policy*, 87: 325–333.

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review*, 104(2): 537–63.

Liang, Jing, Pengfei Liu, Yueming (Lucy) Qiu, Yi David Wang, and Bo Xing. 2018. "Getting More of Something Without Subsidizing It: Impact of Time-of-Use Electricity Pricing on Residential Energy Efficiency and Solar Panel Adoption." *PSN: Renewable Resources & Energy (Topic)*.

Novan, Kevin, Aaron Smith, and Tianxia Zhou. 2022. "Residential Building Codes Do Save Energy: Evidence from Hourly Smart-Meter Data." *The Review of Economics and Statistics*, 104(3): 483–500.

Novan, Kevin, and Aaron Smith. 2018. "The Incentive to Overinvest in Energy Efficiency: Evidence from Hourly Smart-Meter Data." *Journal of the Association of Environmental and Resource Economists*, 5(3): 577–605.

Qiu, Yueming, Matthew E. Kahn, and Bo Xing. 2019. "Quantifying the rebound effects of residential solar panel adoption." *Journal of Environmental Economics and Management*, 96: 310–341.

Rohan, Best, J. Burke Paul, and Nishitateno Shuhei. 2019. "Understanding the Determinants of Rooftop Solar Installation: Evidence from Household Surveys in Australia." *Australia Journal of Agricultural and Resource Economics*, 63(4): 922–939.

US EIA. 2020*a*. "Electric power monthly: March 2020." US Energy Information Administration, Washington, DC.

US EIA. 2020*b*. "Most new utility-scale solar in the United States is being built in the South Atlantic." US Energy Information Administration, Washington, DC.

Zhu, Xing, Lanlan Li, Kaile Zhou, Xiaoling Zhang, and Shanlin Yang. 2018. "A meta-analysis on the price elasticity and income elasticity of residential electricity demand." *Journal of Cleaner Production*, 201: 169–177.

1.7 Figures

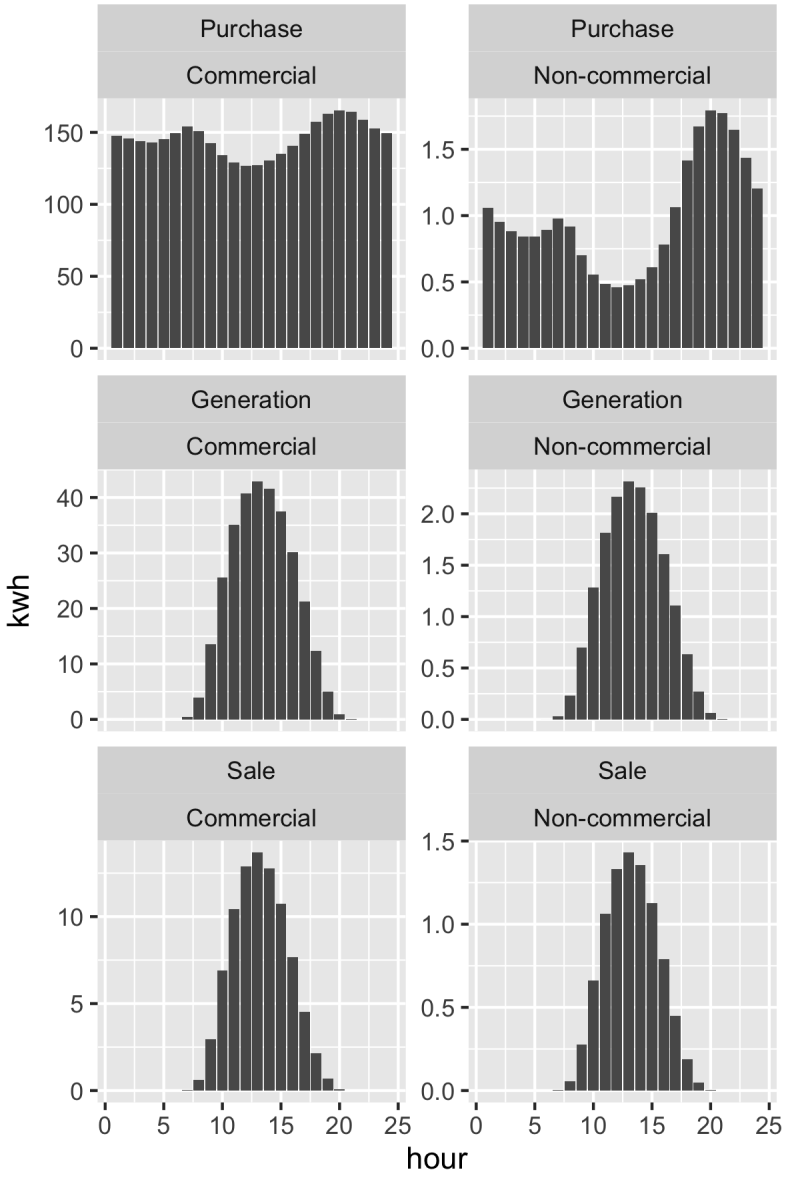
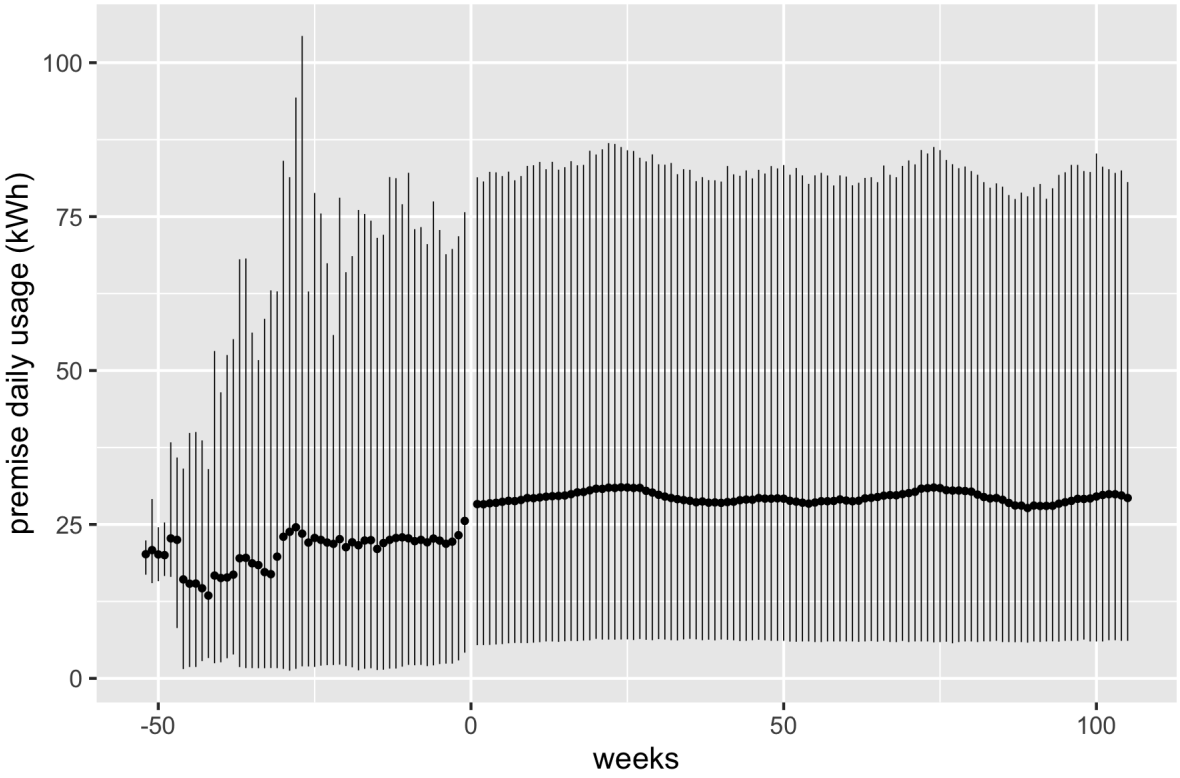
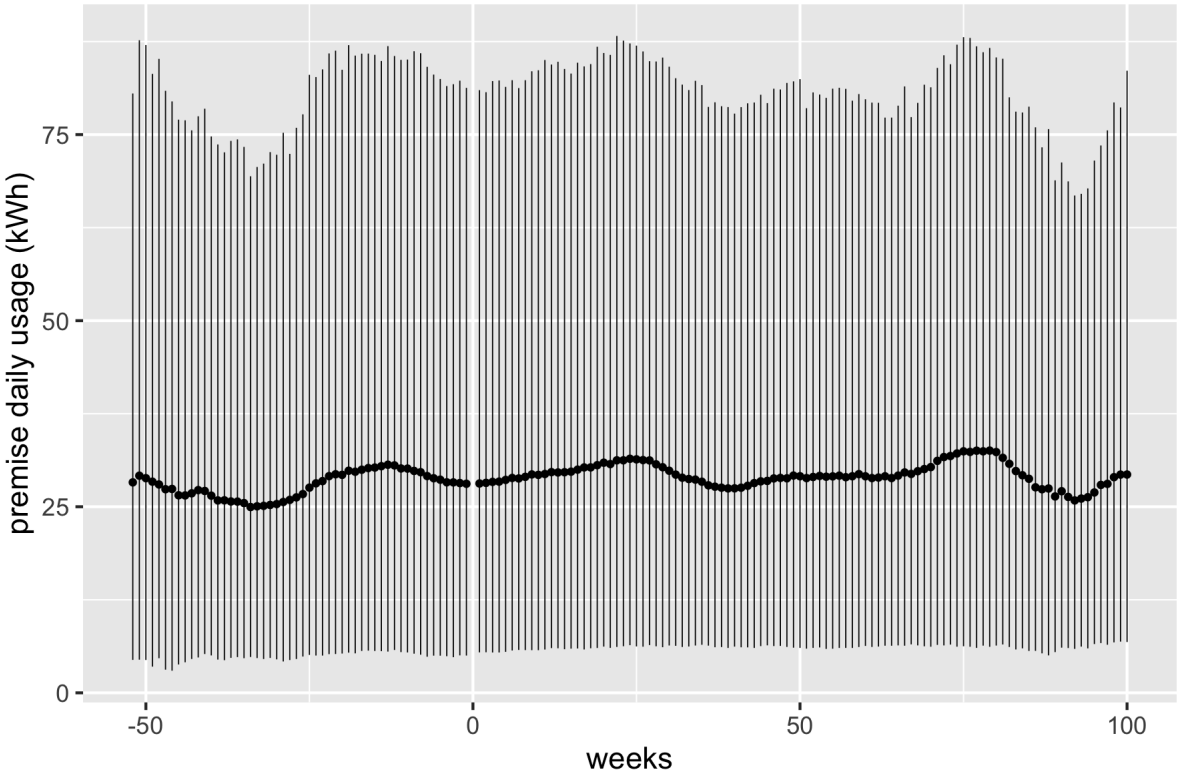


Figure 1.1: Purchase, sale and generation for commercial and non-commercial premises



Notes: Points show the average daily usage (in kWh) across households while the grey bars indicate 2.5% and 97.5% percentiles of average daily usage in a week across all households. We see the average daily usage is discontinuous on the week pre- and post- solar adoption. The 95% quantiles also vary dramatically in the pre-adoption weeks.

Figure 1.2: Consumption by installation weeks before data cleaning



Notes: After data cleaning, the average daily usage (in kWh) across households is continuous on the week pre- and post- solar adoption. The 95% quantiles in the pre-adoption weeks also look more reasonable than before data cleaning.

Figure 1.3: Consumption by installation weeks after data cleaning

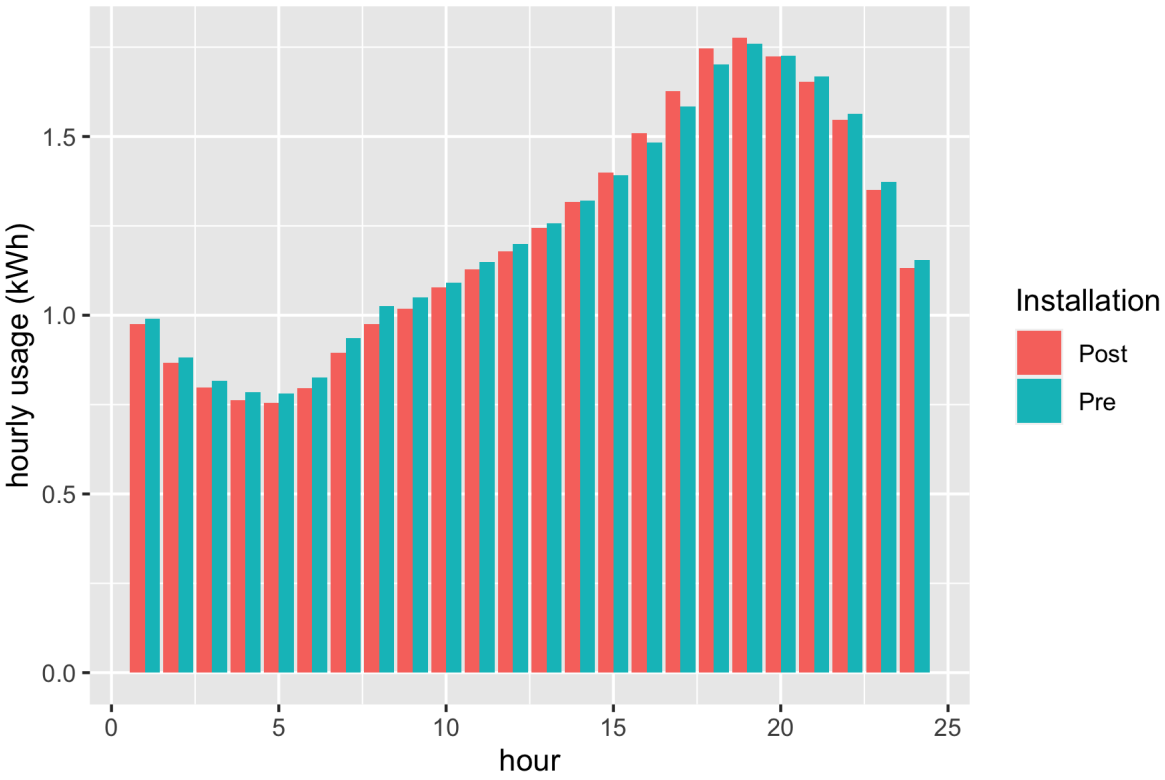


Figure 1.4: Hourly consumption for adopters pre and post solar panels installation

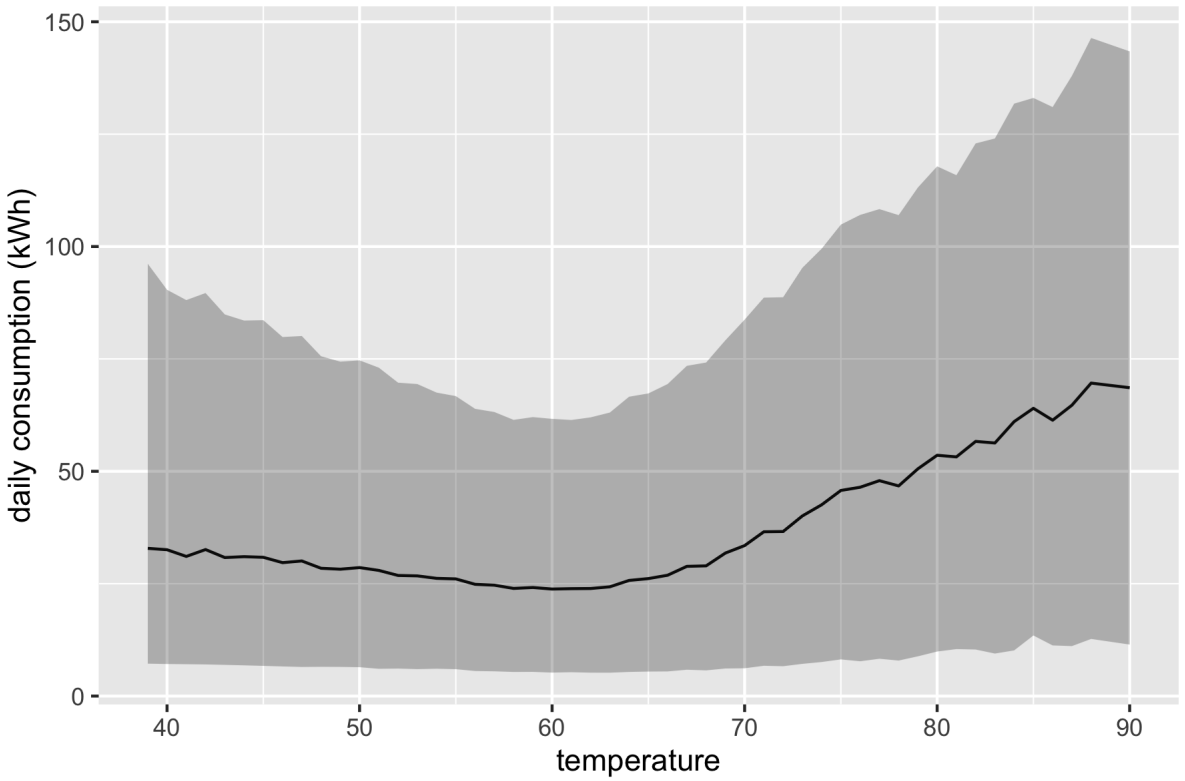


Figure 1.5: Average daily consumption by temperature with 2.5% and 97.5% percentiles

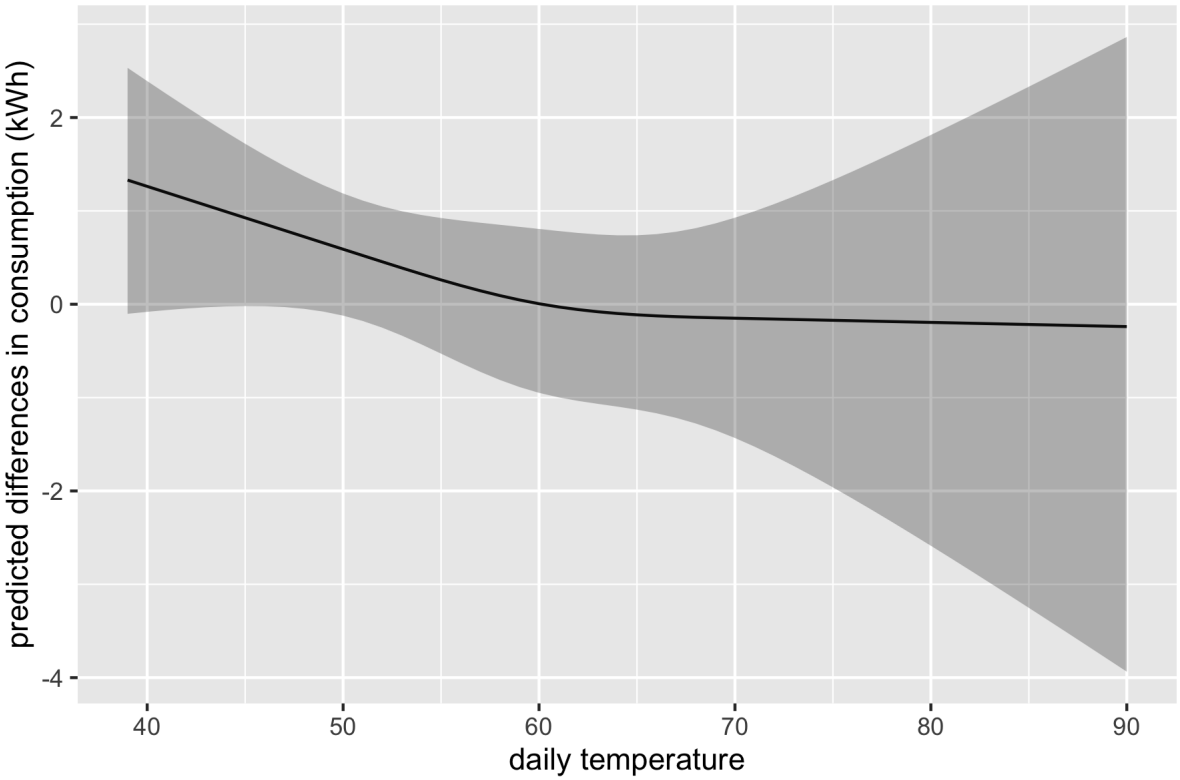


Figure 1.6: Predicted differences in daily consumption pre and post solar panel adoption

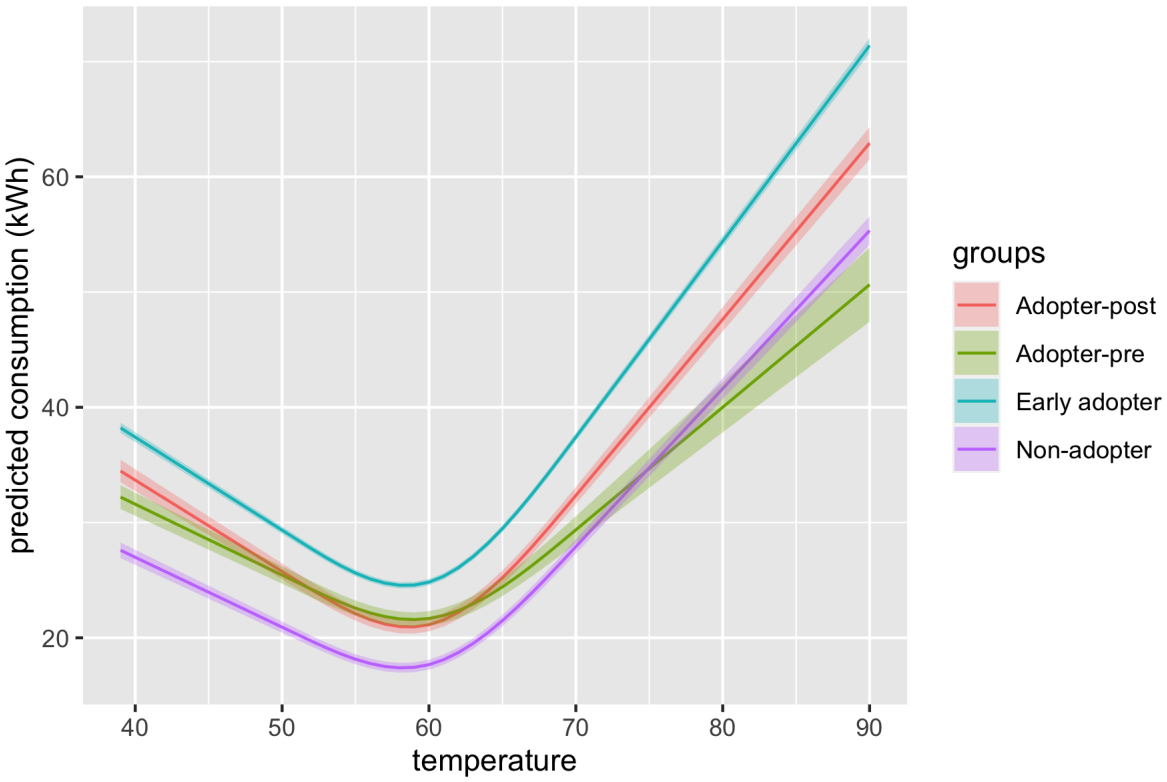


Figure 1.7: Temperature response functions by groups

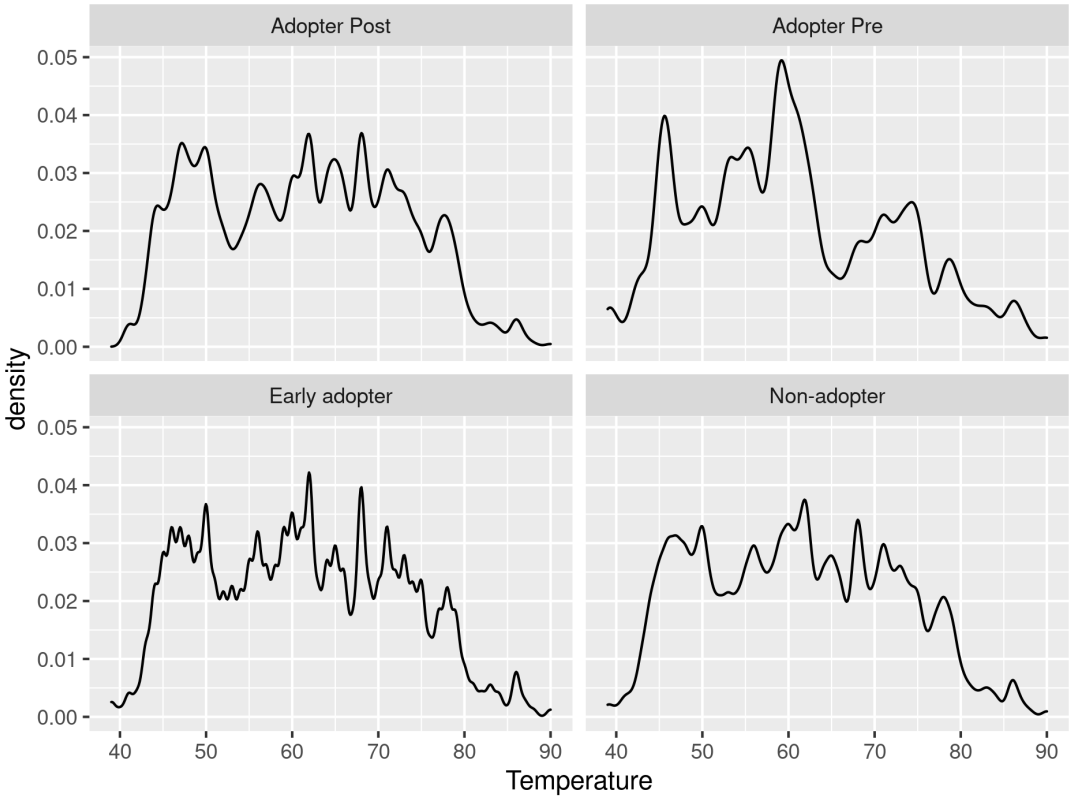


Figure 1.8: Temperature distribution by groups

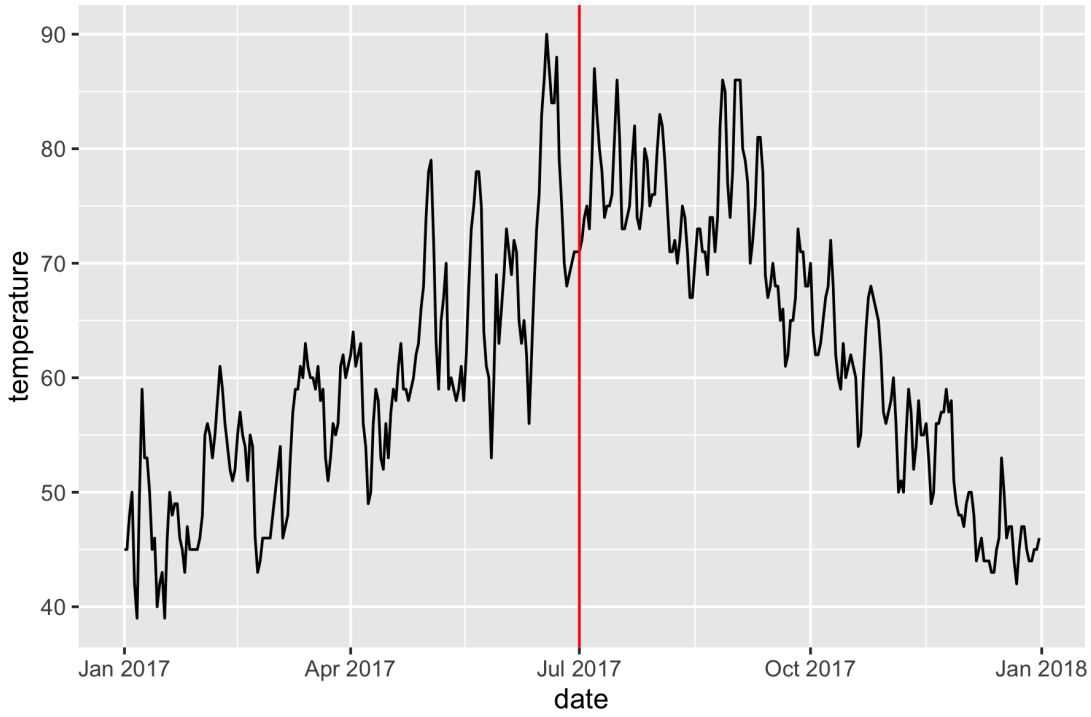


Figure 1.9: Temperature by dates in 2017

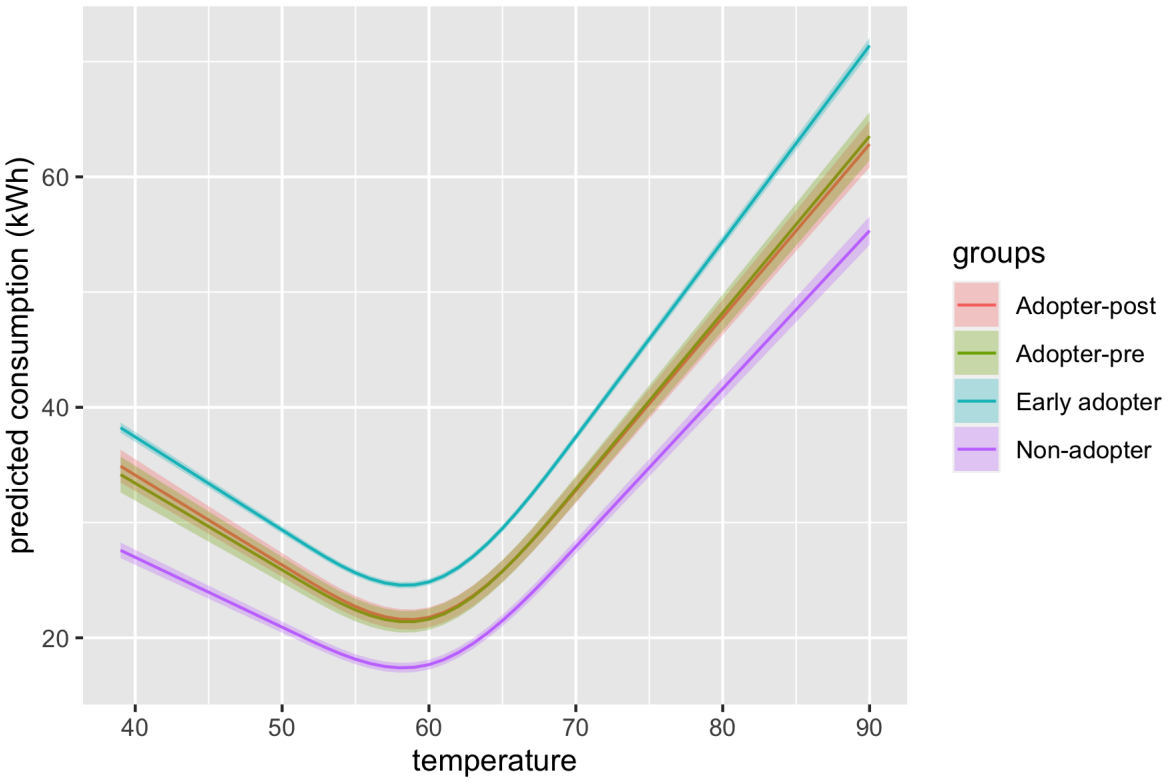
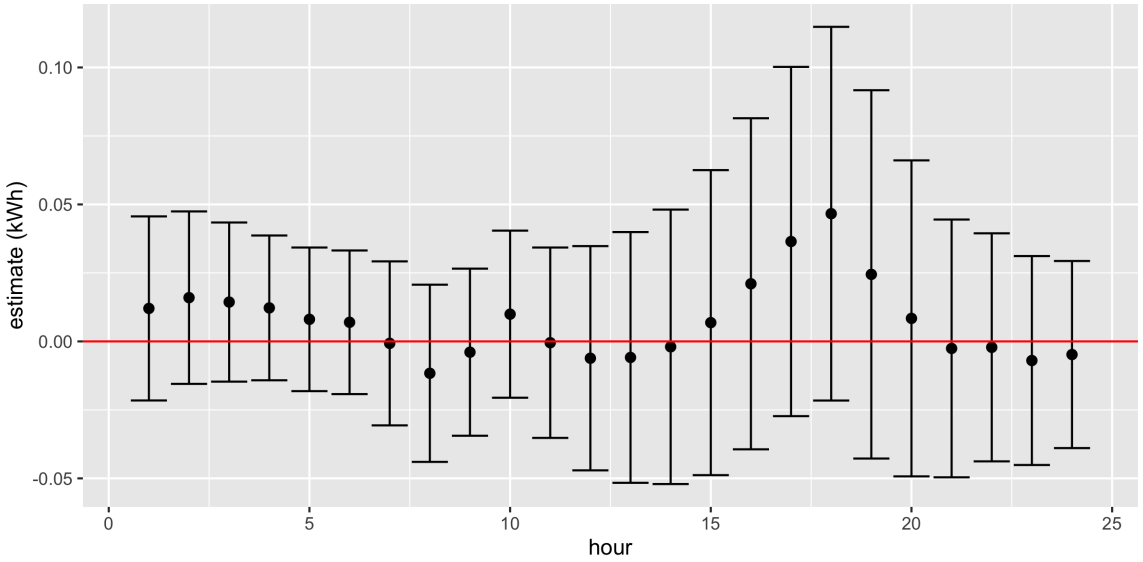


Figure 1.10: Temperature response functions by groups (with adopters install after July)



Notes: We ran the hourly model (equation 1.4) for each hour of a day and saved point estimates and 95% confidence intervals. Rebound effects in different hours are all statistically insignificant at the 5% significance level.

Figure 1.11: Point estimates and 95% C.I. by hours

1.8 Tables

Table 1.1: Summary statistics of electricity usage (in kWh) for households in SMUD area

group	# of premises	variable	mean	median	sd
Panel A: All groups					
Adopter	1659	purchase	22.54	18.06	16.91
		sale	9.22	6.41	10.46
		generation	15.99	13.17	15.71
		consumption	29.31	24.14	20.03
Early adopter	8925	purchase	23.99	19.70	16.95
		sale	10.82	8.44	9.89
		generation	20.57	17.76	14.13
		consumption	33.74	28.46	21.92
Non-adopter	2176	purchase	24.91	20.37	18.07
		consumption	24.91	20.37	18.07
Panel B: Adopters					
Adopter pre	1659	purchase	29.51	23.82	21.10
		consumption	29.51	23.82	21.10
Adopter post	1659	purchase	20.22	16.52	14.55
		sale	12.29	9.95	10.40
		generation	21.31	18.20	14.68
		consumption	29.25	24.24	19.67

Notes: For adopters pre-installation and non-adopters, the solar generation and sale are all 0s.

Table 1.2: Baseline model (equation 1.1) results

	(1)	(2)
Inst	0.175	0.214
	(0.414)	(0.403)
Temp	-0.769 ***	-0.775 ***
	(0.0341)	(0.0358)
Cubic spline (S)	0.00466 ***	0.00451 ***
	(0.00014)	(0.000139)
N obs	9,278,571	1,194,774
N premises	12,760	1,659
TOT1	0.84 %	0.9 %
TOT2	0.73 %	0.9 %
Rebound	0.96 %	1.17 %

Significance levels: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$.
Notes: Column (1) shows results when we run equation 1.1 on the full sample, while column (2) presents results on the adopter-only sample.

Table 1.3: Regression results with interaction terms (equation 1.3)

	(1)
Inst	3.96
	(2.70)
Temp	-0.726 ***
	(0.0467)
Cubic spline (S)	0.00442 ***
	(0.000215)
Inst× Temp	-0.0673
	(0.0523)
Inst× S	0.000122
	(0.000203)
N obs	1,194,774
N premises	1,659

Significance levels: * p < 0.1 ** p < 0.05 *** p < 0.01

Table 1.4: Full sample with alternative fixed effects

	(1)	(2)	(3)	(4)
Inst	0.175	0.526	-0.0429	-0.0436
	(0.414)	(0.333)	(0.504)	(0.303)
Temp	-0.769 ***	-0.767 ***	-0.767 ***	-0.492 ***
	(0.0341)	(0.0337)	(0.0337)	(0.0351)
Cubic spline (S)	0.00466 ***	0.00465 ***	0.00465 ***	0.00378 ***
	(0.00014)	(0.000136)	(0.000136)	(0.000189)
Holiday				1.43 **
				(0.622)
TOT1	0.84%	2.52%	-0.21%	-0.21%
TOT2	0.73%	2.21%	-0.18%	-0.18%
Rebound	0.96%	2.89%	-0.24%	-0.24%
Fixed effects				
Premise	✓	✓		
Year		✓		
Premise × Year			✓	✓
Day of week				✓
Day of month				✓
Month of year				✓

Significance levels: * p < 0.1 ** p < 0.05 *** p < 0.01

Notes: We run our baseline model (equation 1.1) on the full sample. Results in column (1) is the same as the results from our main specification while column (2)–(4) include results from various alternative fixed effects.

Table 1.5: Adopter sample with alternative fixed effects

	(1)	(2)	(3)	(4)
Inst	0.214	0.241	0.128	0.092
	(0.403)	(0.421)	(0.470)	(0.366)
Temp	-0.775 ***	-0.775 ***	-0.775 ***	-0.497 ***
	(0.0358)	(0.036)	(0.036)	(0.0353)
Cubic spline (S)	0.00451 ***	0.00451 ***	0.0045 ***	0.00366 ***
	(0.000139)	(0.00014)	(0.00014)	(0.000181)
Holiday				1.26 **
				(0.581)
TOT1	0.9%	1.01%	0.54%	0.39%
TOT2	0.9%	1.01%	0.54%	0.39%
Rebound	1.17%	1.32%	0.7%	0.51%
Fixed effects				
Premise	✓	✓		
Year		✓		
Premise × Year			✓	✓
Day of week				✓
Day of month				✓
Month of year				✓

Significance levels: * p < 0.1 ** p < 0.05 *** p < 0.01

Notes: We run our baseline model (equation 1.1) on the adopter-only sample. Results in column (1) is the same as the results from our main specification while column (2)–(4) include results from various alternative fixed effects.

Table 1.6: Full sample with alternative knots

	(1)	(2)	(3)
Inst	0.175	0.37	0.614
	(0.414)	(0.403)	(0.406)
Temp	-0.769 ***	-0.844 ***	-0.637 ***
	(0.0341)	(0.036)	(0.035)
Cubic spline (S)	0.00466 ***	0.00194 ***	0.00117 ***
	(0.00014)	(0.0000523)	(0.0000333)
TOT1	0.84%	1.77%	2.94%
TOT2	0.73%	1.55%	2.58%
Rebound	0.96%	2.03%	3.37%

Significance levels: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Notes: We run our baseline model (equation 1.1) on the full sample. Results in column (1) are the same as that of our main specification while column (2)–(3) show results from various alternative cubic spline transformations. We use 10%, 50% and 90% quantiles of temperature distribution in 2017 and 2018 for cubic spline knots in column (2) and 10%, 30%, 50%, 70%, and 90% quantiles for column (3).

Table 1.7: Adopter sample with alternative knots

	(1)	(2)	(3)
Inst	0.214	0.403	0.64
	(0.403)	(0.393)	(0.396)
Temp	-0.775 ***	-0.848 ***	-0.65 ***
	(0.0358)	(0.0378)	(0.0363)
Cubic spline (S)	0.00451 ***	0.00187 ***	0.00113 ***
	(0.000139)	(0.0000532)	(0.000034)
TOT1	0.9%	1.69%	2.69%
TOT2	0.9%	1.69%	2.69%
Rebound	1.17%	2.21%	3.52%

Significance levels: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Notes: We run our baseline model (equation 1.1) on the adopter-only sample. Results in column (1) are the same as that of our main specification while column (2)–(3) show results from various alternative cubic spline transformations. We use 10%, 50% and 90% quantiles of temperature distribution in 2017 and 2018 for cubic spline knots in column (2) and 10%, 30%, 50%, 70%, and 90% quantiles for column (3).

Table 1.8: Baseline models with piecewise linear spline

	(1)	(2)
Inst	0.303	0.333
	(0.385)	(0.378)
Linear spline (W)		
W_1	-0.405 ***	-0.402 ***
	(0.0556)	(0.0551)
W_2	-0.622 ***	-0.65 ***
	(0.0484)	(0.0504)
W_3	1.08 ***	1.01 ***
	(0.0805)	(0.078)
W_4	1.95 ***	1.85 ***
	(0.0841)	(0.081)
TOT1	1.45%	1.4%
TOT2	1.27%	1.4%
Rebound	1.67%	1.83%

Significance levels: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$

Notes: In addition to the cubic spline transformation, we include regression results from the linear spline transformation. We run equation (1.6) on the full sample (results shown in column (1)) and on the adopter-only sample (results shown in column (2)). The linear spline knots are set at 51, 61, and 70°F.

Table 1.9: Baseline models with solar panel sizes

	(1)	(2)
Inst	0.687	0.742
	(0.808)	(0.759)
Temp	-0.776 ***	-0.779 ***
	(0.039)	(0.0381)
Cubic spline (S)	0.00467 ***	0.00451 ***
	(0.000144)	(0.000141)
Size	-0.0234	-0.0243
	(0.0366)	(0.0367)
Gen	0.0133	0.00713
	(0.0201)	(0.0216)
N obs	9,278,571	1,194,774
N premises	12,760	1,659
TOT1	3.29%	3.12%
TOT2	2.88%	3.12%
Rebound	3.78%	4.08%

Significance levels: * p < 0.1 ** p < 0.05 *** p < 0.01

Notes: We run equation (1.7) on the full sample and on the adopter-only sample. Results are shown in column (1) and (2), respectively. The solar system sizes are estimated by the average solar generation for each household in the post-adoption period.

Essay 2

Cooler Summer with Lower Bills: Rebound Effects From AC Upgrades

2.1 Introduction

Residential energy efficiency programs are popular policy instruments that promise to achieve two goals. First, these programs facilitate reducing energy consumption, thus reducing the associated negative externalities such as those from greenhouse gas emissions.[1] Second, these programs help a correct market failure known to the economists as the "energy efficiency gap".[2] However, realized energy savings achieved by residential energy efficiency programs consistently fall short of the ex-ante estimates (Fowlie, Greenstone and Wolfram, 2015). There are several possible explanations. First, the engineering predictions might be simply overinflated for various reasons. The models may fail to capture all real-life complications in heterogeneity among households; the models are made by agents with a stake in

[1] Although the economists agree on that the first-best solutions to address the negative externalities are Pigouvian taxes, such price-based policies often remain politically infeasible. Therefore residential energy efficiency programs are widely viewed as second-best alternative solutions

[2] Ex-ante engineering estimates, such as the well-known Mckinsey curves (McKinsey&Company, 2009), suggest that in many cases consumers are missing out on obvious opportunities to investing in energy-saving technologies that can yield the significant energy savings, sufficient to pay for themselves. A series of important papers address the existence and possible explanations for the energy efficiency gap (Allcott and Greenstone, 2012; Gillingham and Palmer, 2013; Hausman, 1979; Metcalf and Hassett, 1999; Myers, 2020; Rapson, 2014; Sallee, 2014).

the prediction outcome; intentionally biased sample selections can also play a role in that only projects expected to yield higher savings predictions will be carried out, thus causing underperformance on average.

Second, consumers may respond to energy efficiency improvements by using the energy service more intensively. This response, known as a "rebound effect", is typically not included in engineering models. When energy efficiency upgrades happen, the consumer faces lower marginal costs to use the upgraded durables, leading to a series of re-optimizations that create two types of rebound effects in energy use (Borenstein, 2013; Gillingham and Palmer, 2013).[3] The main focus of this chapter is the "direct" rebound effects, which refers to increased energy use from the upgraded durables. When there are increases in energy use resulting from induced changes in consumption of other goods, the increases are called "indirect" rebound effects, which are usually much smaller in magnitude.[4]

Uncovering evidence of a rebound effect, and quantifying how large it may be, typically requires strong assumptions. For example, Dubin, Miedema and Chandran (1986) rely on the assumption that the engineering estimates are correct (Fowlie, Greenstone and Wolfram, 2015).

Davis, Fuchs and Gertler (2014) show in a developing country setting that new AC units can cause electricity consumption to increase. This is clear evidence of a rebound in cooling. However, the researchers only observe the change in energy use. Decomposing this change into energy efficiency improvement and behavioral responses typically requires assumptions. Thus by examining how monthly energy consumption changes pre v.s. post energy efficiency upgrades, the magnitude of the rebound effect (i.e., how much energy use increases because of a behavioral response) cannot be determined without strong assumption.

This chapter estimates the change in the probability of using the AC unit jointly with

[3]We focus on the microeconomic effects. Gillingham, Rapson and Wagner (2016) define a "macroeconomic rebound effect", which is beyond the scope of this chapter.

[4]For example, when a consumer upgrades her AC unit, she might increase her use of her AC because it is less costly to use it for a given effect, which is the direct rebound effect. She might also use the saved dollars from a lower electricity bill on other things like traveling, and the associated increase in her energy use is an indirect rebound effect.

the change in the energy used when it is on. We find evidence of a rebound effect, and we find a large effect. We capture an important rebound margin, which is due to consumers increasing the probability of turning on AC. We do not capture some other rebound margins, such as changes in thermostat settings.

We study the air-conditioning units (AC) Energy Efficiency rebate program implemented by the Sacramento Municipal Utility District (SMUD). Smart meter data provide the opportunity for us to model household-level electricity use with high precision. We employ the mixture of regressions model to uncover the measures of cooling capacity and cooling behavior of households. The variation in the dates on which households received rebates allows us to construct comparable treated and comparison groups, so we use a difference-in-differences (DiD) design to estimate the sizes of the direct effects and the rebound effects in cooling energy use. We have two primary findings. First, the AC Energy Efficiency rebate program is effective in reducing cooling energy use. In the SMUD serving area during 2012–2013, participating households reduced cooling energy use by a considerable amount, averaging 347.10 kWh per household in one summer (1.81 kWh per day in a high-temperature day). Second, we find clear evidence of rebound effects, and the magnitudes of which are significant. The estimated rebound effects from AC units upgrades are 20.61% of direct effects on average.

The chapter proceeds as follows. In section 2.2, we provide an overview of the data we use and an introduction to the AC Energy Efficiency rebate program, followed by our empirical strategy in section 2.3. Next, in section 2.4, we present findings of our main specifications while in section 2.5, results from robustness checks are reported and the sizes and causes of the direct rebound effects are discussed. We conclude in section 2.6.

2.2 Background and Data

To estimate the impact of AC upgrades, we combine household-level smart meter data with information on the participants in SMUD's residential energy efficiency rebate programs. We also acquire temperature data in the Sacramento area, as high temperature is the main cause of usage of AC units. We describe the various datasets in the following sections.

2.2.1 Electricity Consumption

By the beginning of January 2012, each household in the SMUD service territory had a smart meter installed. The smart meters record the hourly aggregate electricity consumed by each household. For the period from January 1, 2012, through December 31, 2013, we observe the hourly consumption data for nearly all of the residential premises in the SMUD service region. In particular, we observe detailed electricity consumption for all these households for two whole summers in 2012 and 2013. Our data encompass both participants and non-participants in the SMUD's AC Energy Efficiency rebate program (explained below). The non-participants are households that did not participate in any energy efficiency programs during our study period, and were randomly selected from the SMUD service area.

2.2.2 The AC Energy Efficiency Rebate Program

We exploit SMUD's AC Energy Efficiency Rebate Program as a quasi-experiment setting and deploy a difference-in-differences (DiD) design. To encourage investments in energy efficiency, SMUD provides its customers with rebates for purchasing new, energy efficient appliances as well as rebates for carrying out energy efficient upgrades for their homes.

This program provides residential customers rebates for installing new air-conditioning (AC) units. To qualify for the rebates, the homeowners must have AC units that meet the EPA's Energy Star standards, and must verify their home's air ducts are properly sealed and that new AC units have the correct refrigerant charge, are properly sized and have adequate air flow. During January 2012 and March 2014, 5,684 single family premises received a rebate for installing a new, energy-efficient AC unit.[5] The data we observe identifies each of these participating households and the dates when rebates were mailed to the participants. While in each instance the rebate mailing dates occurred after the AC or heat pump was installed, the lag between installation and upgrade rebate (as identified by the rebate mailing date) was typically less than one month.[6] Therefore, in our empirical specifications, we treat the

[5] We exclude 130 multi-family premises in this analysis.

[6] To receive the rebate, a participating customer must submit their application for the rebate within 90 days of installing the new unit.

rebate mailing date as accurately representing the date when the upgrade was installed.

In our energy efficiency program data, we do not observe any information on the cost or efficiency ratings of the newly installed AC units. In addition, we do not observe any information on the units that were being replaced — nor whether the premises even had an AC before participating in SMUD's rebate program.[7] While information on the units being installed and replaced would certainly aid in explaining any heterogeneity in the impacts, that information is not necessary for us to produce estimates of the changes in consumption and expenditure.

2.2.3 House and Temperature Information

In addition to the SMUD electricity consumption data, we have linked County Assessor data to each premises in our sample. The Assessor data provide detailed information on the physical characteristics of the homes such as year built, square footage, and the number of stories.

Table 2.1 summarizes selected characteristics of the households in our sample, which include 5,684 participants in the SMUD's AC Energy Efficiency Rebate Program and 3,143 non-participants. The median and mean of 2011 electricity consumption for non-participants are lower than those of participants, perhaps because on average homes of participants have larger homes than non-participants: a larger total number of rooms, and more square feet.

We use National Oceanic and Atmospheric Administration (NOAA) data from the Sacramento International Airport to calculate the daily average temperature in Sacramento on the days of our interest. The choice of an average daily temperature, instead of a specific hour temperature (e.g. 5 pm), is due to the fact that these temperatures are highly correlated (see Table A2 in Novan, Smith and Zhou (2020)). Using either one will yield similar results.

[7] We also do not observe this information for non-participants.

[7] The daily average temperature (in Fahrenheit) is simply the average of hourly temperatures in a day. Each day we have 24 temperatures data for all hours.

2.3 Empirical Strategy

In this section, we first lay out how we define different household groups in our sample. Next, we fit household-level models to estimate the metrics of interest (the total effects on cooling energy, the direct effects, and the rebound effects). Then, we use a differences-in-differences (DiD) design to estimate the differences in these metrics of interest between households that installed more efficient AC units (treated group) in summer 2013 and those that kept using old AC units in the summer of 2013 (compare group). The key underlying assumption of DiD design is that these metrics would have followed a common trend across the treated and comparison groups in the absence of the treatment. We will discuss in detail about what that means along with our econometrics specifications.

2.3.1 Household Groups Definition

We define summer as the period of 5 months in a year: May, June, July, August, September. Figure 2.1 plots the daily average temperature in Sacramento during 2012–2013. The shaded areas on the figure highlight the May–September time window in both years. With a few exceptions, most of the days with daily average temperature above 60 degrees fall within these summer months.

In our data, we observe households divided between AC rebate program participants and non-participants. We further categorize program participants in our sample into five groups according to the dates on which the households received their respective rebates from SMUD:

- Early Participants: households that received rebates before April 30, 2012.
- Summer 2012: households that received rebates between May 1, 2012, and September 30, 2012.
- Treated: households that received rebates between October 1, 2012, and April 30, 2013.
- Summer 2013: households that received rebates between May 1, 2013, and September

30, 2013.

- Late Participants: households that received rebates on or after October 1, 2013.

Figure 2.2 plots the number of households in our data by rebate mailing dates. The shaded areas are our defined summers in 2012 and 2013. We see variations in rebate mailing dates.

In the "Treated" group, we include all the households who got rebated between the two summers. In the "Late Participants" group, we include all the households who got the rebates after the summer of 2013. By construction, it is reasonable to assume that the households in the treated group used an old AC unit during the summer of 2012 and switched to a more efficient model during the summer of 2013, and the households in the late participants and non-participants groups used the same old AC unit through both summers, while the early participants used new AC. Table 2.2 compares the 2011 consumption and house characteristics of treated, early participants and late participants. All variables have close mean and standard deviations for the three groups. We argue that the three participants groups are comparable on observables.

To construct a counterfactual for the treated group, we include early participants, late participants, non-participants as comparison groups in our DiD analysis. The early participants group consists of households who received the rebates before the summer 2012 and thus used new AC units in both summers, while we should expect late participants used the same old AC units in the two summers. We would also expect the cooling functions and behaviors of the treated group converge to those of early participants and diverge from those of late participants in the summer of 2013. Table 2.3 summarizes the numbers of observations in each of the groups. The number of households in each of the buckets is large enough to enable us to draw useful conclusions and insights.

2.3.2 Household-level Model Specification

Assume the daily electricity consumption of household h is a function of the daily average temperature:

$$kWh_h(temp) = \begin{cases} \alpha_h + C_h(temp) + \varepsilon_{1h} & \text{with probability } B_h(temp) \\ \alpha_h + \varepsilon_{2h} & \text{with probability } 1 - B_h(temp) \end{cases} \tag{2.1}$$

We call $B_h(temp)$ the cooling behavior profile of a household and define it as the conditional probability of the household turning on their AC condition on daily average temperature. This function is bounded between 0 and 1 and presumably increases with daily average temperature.

$$B_h(temp) = Pr(\text{AC is on in household h}|temp)$$

Each household is assumed to have baseline electricity consumption level α_h and consumes an additional amount of $C_h(temp)$ when turning on AC. $C_h(temp)$ is thus called the cooling function of the household. It is a function of daily average temperature and is dictated by the household's physical characteristics, especially the efficiency of its AC unit.

This model is estimated separately for each household-year combination so that we not only allow different households to have different baseline consumption, cooling curve, and cooling behavior but also allow them to change from 2012 to 2013.

For all individual households, we assume they will not turn on their AC until the daily average temperature exceeds 60 degrees. Moreover, we define days with average temperature above 85 degrees as outliers.[8] For each household-year, the sample we used to fit the model included only the days with daily average temperature above 60 degrees and below 85. There were 190 such days in 2012 and 191 such days in 2013.

Our first step is to determine the value of α_h. Given daily temperature and daily con-

[8]The daily average temperature exceeded 85°F on 2 days in 2012 and 8 days in 2013, respectively.

sumption data, for each household-year, we regress consumption on temperature and predict the household consumption at 60 degrees.[9] This predicted consumption is our estimated baseline consumption $\hat{\alpha_h}$. We plug this estimated baseline consumption into Equation 2.1, which yields

$$kWh_h(temp) = \begin{cases} \hat{\alpha_h} + C_h(temp) + \varepsilon_{1h} & \text{with probability } B_h(temp) \\ \hat{\alpha_h} + \varepsilon_{2h} & \text{with probability } 1 - B_h(temp) \end{cases} \tag{2.2}$$

Equation (2.2) is a constrained version of a mixture of regressions. In a mixture of regressions model, we assume data come from two (or more) regression models with a mixing probability. In this case, the two regression models are a constant line with the baseline electricity consumption as its intercept and a cooling consumption function that varies with temperature. The mixing probability is the probability of turning on the AC unit. The model is fitted by maximizing a likelihood function of the mixture by an expectation-maximization (EM) algorithm.

Follow Novan and Smith (2018) and Novan, Smith and Zhou (2020), we specify a linear function $C_h(temp)$ in our model.[10] The value of $C_h(temp)$ is fixed to be 0 at 60 degrees, meaning that we assume the household does not use any cooling energy when the daily temperature is 60 degrees. Moreover, we model the probability $B_h(temp)$ to be a non-parametric smooth function.[11]

Figure 2.3 shows results from fitting the model for one of sample premises (id 15508) as an example to illustrate how our model works. On the top panel, dots represent our

[9] Compared with assuming a temperature range and taking the mean or median of the daily consumption to estimate the baseline consumption, this method incorporates more information on the relationship between temperature and consumption, and hence reduces noise introduced by small samples.

[10] We do not strictly follow these two papers by using a simple linear function rather than a piece-wide linear functions. The later functional form is too flexible that would absorb rebound effects

[11] The posterior mixing probabilities are the outputs from each iteration of the algorithm that is used to estimate the mixture of regressions. We then use a smoother that is the local average of the posterior mixing probability. The smoothed function serves as the prior mixing probability for the next iteration and is denoted as $B_h(temp)$ until the algorithm converges.

data points. We first estimate the baseline consumption by regressing daily consumption on temperature and then predicted daily consumption at 60 degrees. This procedure results in two horizontal lines for 2012 and 2013, which represent daily electricity consumption when a household's AC is off. We then use the EM algorithm for each year to find linear cooling functions when AC is on which are two upward sloping lines. Another output from our model is the cooling behaviors $B_h^{2012}(temp)$ and $B_h^{2013}(temp)$ which are depicted in panel (b). Dots represent posteriors from the EM algorithm and the cooling behavior is the weighted average of these posteriors which are non-parametric curves.

2.3.3 Difference in Differences (DiD) Regression

The household-level model also implies that the household's conditional expectation of cooling energy use conditional on temperature is the product of two components, the household's behavior and the cooling ability of the AC unit:

$$CE_h(temp) = E(kWh_h - \hat{\alpha}_h \mid temp) = B_h(temp) \times C_h(temp) \tag{2.3}$$

Since we fit Equation (2.3) for both summers of 2012 and 2013 for all households, we obtained the distinct cooling behavior profiles and cooling curves for each household-year. We denote them as $B_h^{2012}, B_h^{2013}, C_h^{2012}, C_h^{2013}$ for household h.

Total effects

Given a set of summer days with temperature $temp_1, ..., temp_n$, we predict that the total expected amount of kWh household h uses for cooling is TCE_h^{2012} in 2012, and is TCE_h^{2013} in 2013 as follows:[12]

$$TCE_h^{2012} = \sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)$$

[12] In our empirical analysis, the vector of temperatures ($temp_i$) that we use are from the 192 days in 2012 with daily average temperature above 60 degrees.

$$TCE_h^{2013} = \sum_{i=1}^{n} B_h^{2013}(temp_i) \times C_h^{2013}(temp_i)$$

Then the total change in cooling energy use is

$$\begin{aligned} TC_h &= TCE_h^{2013} - TCE_h^{2012} \\ &= \sum_{i=1}^{n} [B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)] \end{aligned}$$

and the total effects from the treatment are defined as

$$TS_h = E(TC_h|s = Treated) - E(TC_h|s = Comparison)$$

where s is an indicator for treatment group assignment. Recall that we define the treated group to be households that received rebates between October 1, 2012, and April 30, 2013. The comparison groups consist of early participants, late participants and non-participants. In our analysis, we contrast the treated group with different combinations of the three "non-treated" groups.

To capture the total effects in a regression DD framework, we can employ the following equation:

$$\begin{aligned} TCE_{hs}^t = Y_{hs}^t = \beta_0 + \beta_1 \cdot 1\{s = Treated\} + \beta_2 \cdot 1\{t = 2013\} \\ + \gamma \cdot (1\{s = Treated\} \cdot 1\{t = 2013\}) + \varepsilon_{hs}^t \end{aligned}$$

where h is the household subscript, s is the treatment subscript, with value equal to "Treated" or "Comparison," t is the time subscript with value equals to "2012" or "2013."

If ε_{hs}^t have different variances for different t, we estimate the first difference form:

$$TC_{hs} = \Delta Y_{hs} = \beta_2 + \gamma \cdot 1\{s = Treated\} + \Delta\varepsilon_{hs} \tag{2.4}$$

where we define the left hand side (LHS) variable in Equation (2.4) as

$$\Delta Y_{hs} = TC_{hs} = \sum_{i=1}^{n} B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - \sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)$$

we have $\hat{\gamma}$ as an estimator of the total effects

$$\begin{aligned}\hat{\gamma} =& E\{\sum_{i=1}^{n}[B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)]|s = Treated\} \\ & - E\{\sum_{i=1}^{n}[B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)]|s = Comparison\}\end{aligned}$$

The underlying assumption of such a DiD regression specification is that the total cooling electricity use (TCE_{hs}) will have common trends in the treated households and the comparison households, in the absence of the treatment. This assumption is reasonable because of two facts: (1) treated households, late participants and non-participants would have used the same old AC units in the absence of the treatment (i.e., AC upgrade) in both summers while early participants used the same new AC units. Observables are comparable within participants and the differences in characteristics between participants and non-participants are not dramatic. (2) we expect that the cooling behaviors for all household do not change in the absence of the treatment. This also aligns with our results in section 2.4.1.[13]

Direct effects

Conceptually, direct effects are the effects from only the efficiency improvement, assuming no behavioral changes (i.e., no changes in turning off or on the AC units under the same temperature across years). With our model, direct effects differ from the total effects in that direct effects refer to changes in energy use presuming there are no changes in the probability

[13]One potential threat to the parallel trend assumption is how people chose whether and when to participate may be correlated with their cooling functions or behaviors. However, in Figure 2.5 and Figure 2.7, regardless of participation or not and when to participate, the early participants, the late participants and the non-participants display similar trends in the changes in cooling and behavior functions.

of turning on the AC. We can define direct effects as

$$
\begin{aligned}
DS_h =& E\{\sum_{i=1}^{n} B_h^{2012}(temp_i) \times [C_h^{2013}(temp_i) - C_h^{2012}(temp_i)] | s = Treated\} \\
& - E\{\sum_{i=1}^{n} B_h^{2012}(temp_i) \times [C_h^{2013}(temp_i) - C_h^{2012}(temp_i)] | s = Comparison\} \\
=& E\{\sum_{i=1}^{n} [B_h^{2012}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)] | s = Treated\} \\
& - E\{\sum_{i=1}^{n} [B_h^{2012}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)] | s = Comparison\}
\end{aligned}
$$

This definition relies on two key assumptions. First, we assume that unchanged household behavior translates to unchanged $B_h(temp)$. Accordingly, we use the estimates of the cooling behavior response for 2012 ($B_h^{2012}(temp)$) combined with cooling functions we estimated for both 2012 and 2013 to estimate the direct effects. This identification assumption seems reasonable in cases where average temperature explains cooling electricity use pretty well. Our data also support this assumption. For comparison groups where we expect no behavioral changes, estimated $B_h^{2012}(temp)$ and estimated $B_h^{2013}(temp)$ are quite close as we will show in the results.

The second assumption is the common trends assumption on which the DD regression relies. We assume that the cooling functions of the AC units will have common trends in the treated households and the comparison households, in the absence of the treatment. This assumption is reasonable because both treated households and comparison households will use similar AC units in the absence of the treatment and the two groups of households are comparable in other household characteristics.

Using this of direct effects, we can easily estimate direct effects using DiD regression by modifying Equation (2.4) to

$$\Delta Y_{hs}^{*} = \beta_2 + \gamma \cdot 1\{s = Treated\} + \Delta\varepsilon_{hs} \tag{2.5}$$

where we define

$$\Delta Y^*_{hs} = \sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2013}(temp_i) - \sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)$$

Rebound effects

We define rebound effects as the differences between total effects and direct effects. Thus, under our definition of direct effects above, the rebound effects are defined as

$$\begin{aligned} RE_h =& TS_h - DS_h \\ =& E\{\sum_{i=1}^{n}[B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2013}(temp_i)]|s = Treated\} \\ & - E\{\sum_{i=1}^{n}[B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2013}(temp_i)]|s = Comparison\} \end{aligned}$$

we estimate the rebound effect using DiD regression by modifying Equation (2.4) to

$$\Delta Y^{**}_{hs} = \beta_2 + \gamma \cdot 1\{s = Treated\} + \Delta\varepsilon_{hs} \tag{2.6}$$

with

$$\Delta Y^{**}_{hs} = \sum_{i=1}^{n} B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - \sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2013}(temp_i)$$

2.4 Results

In this section, we display our results as follows. First, we look at our predicted cooling functions and cooling behaviors, which also give us insights on what was going on in the households and what caused the rebound effects. Next, we discuss our DiD regression results and the sizes and significance of the total effects, the direct effects, and the rebound effects.

2.4.1 Cooling Functions and Behavioral Changes

Our estimated cooling functions and cooling behaviors are both functions of daily temperature ($C_h(temp)$ is linear in temperature while $B_h(temp)$ is non-parametric). We average these functions across households and plot them by household groups.

Figure 2.4 shows the average cooling curves in 2012 and 2013 for each of the household groups. Across the groups, we see that the "Treated" and "Late Participants" groups had similar average cooling lines (similar slopes for linear functions) in 2012, and the "Treated" and "Early Participants" groups had similar average cooling lines in 2013. These findings align with our perception that the "Treated" and "Late Participants" groups both used old AC units in 2012 and the "Treated" and "Early Participants" groups both used new AC units in 2013. Within groups, "Treated" households have lower average cooling curves in 2013 than in 2012, showing that the AC upgrades indeed improved cooling efficiency, while other groups show little or no differences between the two years.

Figure 2.5 plots the differences (2013 predicted cooling energy use minus 2012 predicted cooling energy use under every temperature) between the pairs of cooling lines, along with pointwise 95% confidence intervals generated by the two-sample t-tests. These plots confirm that AC upgrading significantly improves energy efficiency in both economic and statistical sense.

Notice that the late participants have a significantly lower cooling function in 2013 than that in 2012. This could reflect the fact that there were lags between the AC upgrade installation date and the rebate date, although in our main analysis we assume these two dates are the same. In order to test how the lags would affect our estimates of total effects, direct effects and rebound effects, in Appendix 2.A.2, we put off the rebate dates for treated group and late participants to October 15 and November 1. That is, we test two specifications in which (1) the treated group are households who received rebates between October 15, 2012 and April 30, 2013, and late participants are households who received rebates on or after October 15th, 2013; (2) the treated group received rebates between November 1, 2012 and

April 30, 2013, while late participants received rebates on or after November 1st, 2013. The results show no dramatic changes when we redefine groups.

Figure 2.6 shows the average cooling behaviors in 2012 and 2013 for each of the household groups. In the "Treated" groups, average households have higher probabilities of turning on their ACs in 2013 than in 2012, starting from around 67 degrees, showing behavioral evidence of the rebound effects. In contrast, the gaps are much smaller between average cooling behaviors in 2013 vs 2012, for households in the "Late Participants", "Early Participants" and "Non Participants" groups. Notice that, under around 67 degrees, the probabilities of turning on ACs are higher in 2012 than 2013 for all groups. One potential explanation is we restrict every household to potentially turn on its AC only for temperatures of 60°F and above. However, in the real world, households might choose to turn on their ACs on days with average daily temperatures below 60°F making it hard to separate cooling energy use and non-cooling energy use when temperature is mild. As a result, our measures of cooling behavior under 70 degrees are not well defined. When temperature increases, the difference between cooling and non-cooling consumption becomes easier to detect by the EM algorithm.

Figure 2.7 plots the differences between the pairs of cooling behaviors showed in Figure 2.6, along with point-wise confidence intervals generated by the two-sample t-test. It suggests that at daily average temperatures above 70 degrees, the behavioral changes are often significant for the "Treated" groups, but not quite so for the other groups. The pattern of the differences is also interesting. The probability increase for the "Treated" group takes off from 67 degrees and then almost stabilizes after 75 degrees.

Figure 2.8 depicts estimates of total effects, direct effects, and rebound effects, using the measures that were defined in Section 2.3.3, which are $B_h^{13}(temp) \times C_h^{2013}(temp) - B_h^{2012}(temp) \times C_h^{2012}(temp)$, $B_h^{2012}(temp) \times C_h^{2013}(temp) - B_h^{2012}(temp) \times C_h^{2012}(temp)$, and $B_h^{2013}(temp) \times C_h^{2013}(temp) - B_h^{2012}(temp) \times C_h^{2013}(temp)$, respectively. We calculate these metrics for each household and then calculate the means along with 95% confidence intervals based on t-distributions. The plots show statistically significant total effects, direct effects,

and rebound effects for the treated group. Moreover, the rebound effects increase when temperature is high. This is a result from our linear cooling functions. As shown in Figure 2.5, the higher the daily average temperature is, the more cooling consumption is conserved after upgrading AC units. Meanwhile, we observe no significant rebound effects from other groups.

Notice that late participants exhibit significant and slightly negative total and direct effects. Again, this is because some households who upgraded their AC units in 2013 summer but apply for rebates after Oct 1st, 2013 were categorized as late participants. The early participants and non-participants had no significant change in consumption across the two years. In order to quantify total, direct and rebound effects, we move to DiD analysis in the next section.

2.4.2 Difference in Differences (DiD) Regression Results

Figure 2.9 shows the distribution of the predicted total changes in cooling energy use ($\widehat{TC_h} = \widehat{TCE_h^{2013}} - \widehat{TCE_h^{2012}}$). Negative total changes correspond to decreases in total cooling energy use from 2012 to 2013. In the densities for each of the other groups, the total changes center around 0 while the treated group's density peaks at a negative value. The households in the treated group are more likely to have decreased their cooling energy use, compared with other groups.

In Table 2.4 we jointly present total effects, direct effects and rebound effects. These estimates are $\hat{\gamma}$ from Equation (2.4), (2.5), and (2.6). Detailed regression results for the three equations can be found in Appendix 2.A.1. We run regression on four different samples in which the treated group is compared with combinations of Early Participants, Late Participants, and Non-participants.

The total effects for different comparison groups are estimated to be 347.10, 383.62, 442.95, and 420.86 kWh, in one summer, respectively.[14] The direct effects are estimated to

[14]To put these numbers into perspective, recall that we use 192 higher-than-60 days in 2012 as our temperature vector in generating the predictions. Thus these total effects correspond to 1.81, 2.00, 2.31, and 2.19 kWh per day. If we apply a flat rate of \$0.128 per kWh (see Novan and Smith (2018)), households will save \$6.95, \$7.68, \$8.87, and \$8.41 per month (30 days) during summer.

be 437.21, 456.99, 497.02, and 482.12 kWh. It is not surprising to see that comparing non-participants with the treated group yields the greatest estimate of total and direct effects, as the non-participants on average have slightly smaller houses and historically consume less electricity than AC rebate program participants.

On average, across the four comparison groups we would expect to see rebound effects of 90.11, 73.38, 54.07, and 61.26 kWh over the 192 days of summer 2012. These numbers correspond to 20.61%, 16.01%, 10.88%, and 12.71% of the direct effects when we contrast the treated group with the various comparison groups.

2.4.3 Social cost savings and private gains

Gains from the SMUD AC rebate program are two-fold. On the one hand, AC upgrading increases the households' cooling efficiency during warm months. This could reduce the cost of producing energy (i.e., private generation costs and external costs). On the other hand, households gain welfare by paying less bills (decreases in total expected cooling energy use. i.e., the total effects on average across households are negative) and by consuming more electricity but enjoying cooler in-house temperatures (positive average rebound effects across households).

Follow Novan and Smith (2018), we define the change in the social cost of household h in day i as:

$$\Delta Social\ cost_h(temp_i) = (\rho_i + \mu) \cdot [B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)] \tag{2.7}$$

where ρ_i and μ denote the marginal private cost of supplying electricity and the marginal external cost of supplying electricity. For ρ_i, we use the average locational marginal price (LMP) during 6 p.m. through 9 p.m. in day i, when the majority of the cooling savings occur (see Figure 2 in Novan and Smith (2018)).[15] Given that natural gas fired generators

[15] In addition to the savings from the marginal private cost of supplying electricity and the marginal external cost of supplying electricity, the SMUD's AC upgrading program may also provide avoided costs in generation capacity. However, as estimated in Novan and Smith (2018), the avoided capacity costs from a typical

are on the margin the vast majority of time (particularly during the period being studied), it is sensible to focus on avoided costs in CO_2 generation for the estimation of the external pollution cost per kWh μ. The CO2 rate for a typical combined cycle natural gas generator is roughly 0.44 tons CO2/MWh. If we assume that a ton of CO2 has an external cost of \$50/ton, then that would imply that each kWh of electricity avoided would reduce the external costs from CO2 emissions by 2.2 cents per kWh:

$$\$50/ton \cdot 0.44 tons/MWh \cdot 1MWh/1000kWh = \$0.022/kWh$$

Follow this method and results (i.e., total effects in equation 2.4) from our main specification, and summing up social cost changes over date i (recall that there were 192 such days in 2012), the average social costs saving across households is estimated at \$31.58 per summer.

To measure private welfare gains, we first focus on the household's savings in electricity bills. For a household h, savings from upgrading AC units are the product of the tiered price p_h and the changes in total consumption across warm months (i.e., i = 1,...,n):

$$\Delta Private\ cost_h = \sum_{i=1}^{n} p_h \cdot [B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)] \tag{2.8}$$

SMUD provides two rate categories. For households which consume less than 700 kWh per month during June 1st to September 30th, the rate is 9.89 ¢/kWh, and 18.03 ¢/kWh otherwise. We determine a household's rate tier p_h by calculating their 2012 monthly average electricity consumption during 2012. As shown in Figure 2.10, most households fall in the second tier and pay 18.03 ¢/kWh during the 2012 summer season. On average, households pay \$94.90 less in one summer after upgrading their AC units. Another important channel for

household in the sample is on the order of \$0.20 per month, which is almost negligible compared to other social cost savings and private gains. Hence, we do not include the avoided generation capacity costs in our calculation.

private welfare gains is the rebound effects. Households sacrifice some of their direct effects (DS_h in the previous section) by turning on their AC more often during warm months to gain cooler in-house temperatures. The welfare gains for household h from the rebound effects can be estimated as:

$$\Delta Private\ gain_h = \sum_{i=1}^{n} p_h \cdot [B_h^{2013}(temp_i) \times C_h^{2013}(temp_i) - B_h^{2012}(temp_i) \times C_h^{2013}(temp_i)] \tag{2.9}$$

notations are the same as in equation 2.8 except for we are now multiply tiered prices by rebound effects RE_h. The estimated gains on average is \$6.80 in a summer. Ignoring this private welfare gain will understate the benefits of the SMUD AC energy efficiency rebate program.

2.5 Robustness Checks

2.5.1 Negative predicted cooling energy use

In our results, not all households have positive predicted cooling energy use. That is, one or more of the four estimates that are used to construct total effects, direct effects, and rebound effects is negative.[16] By construction, the coefficients representing cooling behavior are positive numbers from 0 to 1. Therefore, the only possible reason for getting negative cooling energy use is through cooling functions that are negative in $temp_i$. This could happen if households tend to leave their residence during hot degree days, resulting in a negative relationship between temperature and electricity consumption.

Table 2.5 shows results after excluding households with negative predicted cooling energy use. We do not see obvious changes in our estimates of total effects, direct effects and rebound effects. Detailed regressions results for Equation (2.4), (2.5), and (2.6) are included

[16] Some households have negative estimated values for $\sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2012}(temp_i)$; $\sum_{i=1}^{n} B_h^{2012}(temp_i) \times C_h^{2013}(temp_i)$; $\sum_{i=1}^{n} B_h^{2013}(temp_i) \times C_h^{2012}(temp_i)$; $\sum_{i=1}^{n} B_h^{2013}(temp_i) \times C_h^{2013}(temp_i)$, where $temp_i$ are temperatures above 60 degrees in 2012. This reflects negative estimated values for $C_h^{2012}(temp_i)$ as discussed in the text.

in Appendix 2.A.3.

2.5.2 Parametric cooling behaviors

De Veaux (1989) introduced the EM algorithm to find solutions to mixture of regressions. In the original EM algorithm, the mixing probability is a constant function (the mean of the posterior probability after each iteration), which is not ideal for situations such as ours when the probability of turning on the AC is expected to be an increasing function of temperature. Nevertheless, the constant mixing probability is easy to understand and implement, and provides useful insights.

Young and Hunter (2010) improved upon the classic EM algorithm by allowing the mixing probability to be approximated nonparametrically in terms of the predictor. They used local polynomial regressions to smooth the posterior probabilities in each iteration. Our main specification is a special case of their method where the degree of the local polynomial function is 0 (i.e., local weighted average). Here we consider another specification in which the mixing probability is a linear function of temperatures. One caveat of using linear functions is that the estimated probabilities are not guaranteed to be bounded in the interval $[0, 1]$. This, however, can be easily fixed by forcing all estimates to fall between 0 and 1.

Table 2.6 summarizes regression results, using different functions to estimate the mixing probability. We include our main results from Table 2.4 for comparison. Overall, more flexible functions give higher rebound effects. For example, in column (1) where we run DiD regressions on treated and late participants, the local weighted average function produces the highest rebound effects among the three methods, while if we use a constant function, the rebound effect is no longer statistically significant.

Recall that the rebound effect is defined as $\sum_{i=1}^{n} C_h^{2013}(temp_i)[B_h^{2013}(temp_i) - B_h^{2012}(temp_i)]$ for each household h. The difference between cooling behaviors across the two summers is critical for identifying the rebound effects. In Appendix Figure 2.B.4, we can clearly see that for the treated group, the constant function smooths out the difference in cooling probabilities across the 2013 summer and the 2012 summer, while the difference increases from

negative to positive when we use the linear mixing probability. Additionally, the linear cooling function $C_h^{2013}(temp_i)$ is also increasing in temperatures. Therefore, mixing probability functions that yield positive probability differences in higher temperatures tend to produce higher rebound effects.

The linear mixing probability method also produces the highest estimates of total and direct effects comparing with those estimated under constant and local weighted average mixing probabilities. Given that the cooling functions are almost identical across all mixing probability smoothing functions (Appendix Figure 2.B.1 and 2.B.2), it is not surprising that smoother behavior functions such as constant and local weighted average mixing probabilities would give lower total and direct effects.

These results confirm that the standard mixture regression algorithm with a constant mixing probability may not be the appropriate method for our study. Nevertheless, the constant mixing probability approach also finds that households in the treated group have a significantly higher probability of turning on their ACs on after upgrading.

2.5.3 Span of local regression

In Section 2.3.2, we describe how we approximate the cooling behaviors using local averages of the posterior mixing probability. One key parameter for such a procedure is the span, which is the bandwidth that we use for local averages. In our setting, we apply the default setting in the R *loess* package to set the span as 0.75, which means for each data point, the nearest 75% data points are used to calculate the weighted local average. Here we present regression results from two other specifications where the span is set to be 0.6 and 0.9.[17]

Table 2.7 presents results with alternative spans, we again include our main results (with span equals 0.75) as a benchmark. Overall, conclusions from our main specification holds when we switch span from 0.75 to 0.6 and 0.9, although the significance levels and magnitudes of rebound effects increases as we narrowing bandwidths. The same intuition in the previous

[17]The optimal span for each household-year data can be found by cross-validations. However, cross-validations are computationally expensive to implement given our sample size and will not necessarily produce a clearer pattern in cooling behaviors than our current method when averaging across all households. Therefore, we proceed with our current method and use the same span for every household-year.

section also applies here. When we use a span of 0.9, we incorporate more data points to approximate behaviors at every temperature. As a result, the cooling behavior functions ($B(temp)$) become smoother than those estimated under a span of 0.6 or 0.75. Smoother mixing probabilities yield smaller rebound effects. In Appendix 2.B.2, the cooling functions for the different groups are almost identical under alternative spans. The span has more observable impacts on the cooling behaviors.

2.6 Conclusion

In this chapter, we estimate the total, direct and, rebound effects of the AC Energy Efficiency rebate program implemented in the SMUD serving area. The average participating household saved 347.10 kWh in one summer (1.81 kWh per day when the temperature is above 60°F).[18] Average direct effects and rebound effects are estimated to be 437.21 kWh and 90.11 kWh in one summer, which translate to a 20.81% direct rebound effect from the AC upgrades. Cooling curve estimation shows that the AC upgrades program is effective in improving the electricity efficiency. Cooling behavior estimation shows that households turn on their new AC more often, which leads to the rebound effects. We examine alternative specifications regarding our sample, functions of mixing probabilities, and local polynomial regression spans. Our main conclusions do not change and are robust under various settings.

Our analysis is not intended to be a comprehensive evaluation of the SMUD's AC Energy Efficiency rebate program. For that purpose, we would need to also model program participation. [19] Rather, it serves as an estimation of the treatment effects on treated and the main goal is to shed light on the sizes of rebound effects in the context of energy efficiency programs.

Our methodology has much potential in the future. As utilities more and more commonly use smart meters, more future research can be done to validate our results. It also has the potential to help target subgroups of households and increase the effectiveness of energy

[18]The average daily saving is calculated by diving 347.10 kWh by 192 days.
[19]See works of Hartman (1988); Mills and Schleich (2010); Train (1988)

efficiency programs or other policies such as information provision.

References

Allcott, Hunt, and Michael Greenstone. 2012. "Is There an Energy Efficiency Gap?" *The Journal of Economic Perspectives*, 26(1): 3–28.

Borenstein, Severin. 2013. "A Microeconomic Framework for Evaluating Energy Efficiency Rebound And Some Implications." National Bureau of Economic Research.

Davis, Lucas W, Alan Fuchs, and Paul Gertler. 2014. "Cash for coolers: evaluating a large-scale appliance replacement program in Mexico." *American Economic Journal: Economic Policy*, 6(4): 207–238.

De Veaux, Richard D. 1989. "Mixtures of linear regressions." *Computational Statistics & Data Analysis*, 8(3): 227–245.

Dubin, Jeffrey A, Allen K Miedema, and Ram V Chandran. 1986. "Price effects of energy-efficient technologies: a study of residential demand for heating and cooling." *The RAND Journal of Economics*, 17(3): 310–325.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2015. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." National Bureau of Economic Research WP 21331.

Gillingham, Kenneth, and Karen Palmer. 2013. "Bridging the energy efficiency gap: Insights for policy from economic theory and empirical analysis." Resources for the Future Discussion Paper 13-02.

Gillingham, Kenneth, David Rapson, and Gernot Wagner. 2016. "The Rebound Effect and Energy Efficiency Policy." *Review of Environmental Economics and Policy*, 10(1): 68–88.

Hartman, Raymon S. 1988. "Self-selection bias in the evolution of voluntary energy conservation programs." *The Review of Economics and Statistics*, 70(3): 448–458.

Hausman, Jerry A. 1979. "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables." *The Bell Journal of Economics*, 10(1): 33–54.

McKinsey&Company. 2009. "Unlocking Energy Efficiency in the U.S. Economy."

Metcalf, Gilbert E, and Kevin A Hassett. 1999. "Measuring the energy savings from home improvement investments: evidence from monthly billing data." *Review of economics and statistics*, 81(3): 516–528.

Mills, Bradford F, and Joachim Schleich. 2010. "Why don't households see the light?: Explaining the diffusion of compact fluorescent lamps." *Resource and Energy Economics*, 32(3): 363–378.

Myers, Erica. 2020. "Asymmetric information in residential rental markets: Implications for the energy efficiency gap." *Journal of Public Economics*, 190: 104251.

Novan, Kevin, Aaron Smith, and Tianxia Zhou. 2020. "Residential Building Codes Do Save Energy: Evidence from Hourly Smart-Meter Data." *Working Paper*.

Novan, Kevin, and Aaron Smith. 2018. "The Incentive to Overinvest in Energy Efficiency: Evidence from Hourly Smart-Meter Data." *Journal of the Association of Environmental and Resource Economists*, 5(3): 577–605.

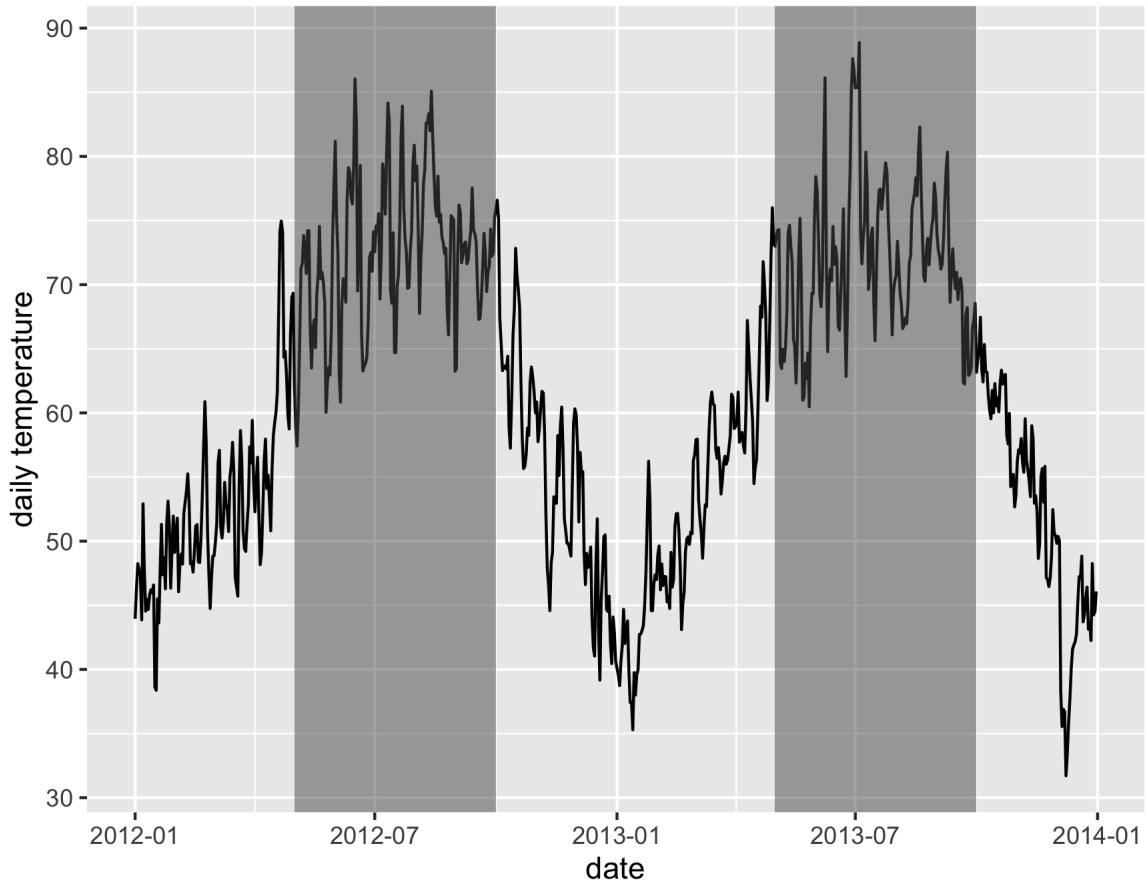
Rapson, David. 2014. "Durable goods and long-run electricity demand: Evidence from air conditioner purchase behavior." *Journal of Environmental Economics and Management*, 68(1): 141–160.

Sallee, James M. 2014. "Rational inattention and energy efficiency." *The Journal of Law and Economics*, 57(3): 781–820.

Train, Kenneth E. 1988. "Incentives for energy conservation in the commercial and industrial sectors." *The Energy Journal*, 9(3): 113–128.

Young, Derek S, and David R Hunter. 2010. "Mixtures of regressions with predictor-dependent mixing proportions." *Computational Statistics & Data Analysis*, 54(10): 2253–2266.

2.7 Figures



Note: We observe the hourly temperatures from a NOAA station at the Sacramento International Airport and calculate the daily temperatures by averaging hourly temperatures.

Figure 2.1: Average daily temperature (°F) in Sacramento during 2012-2013

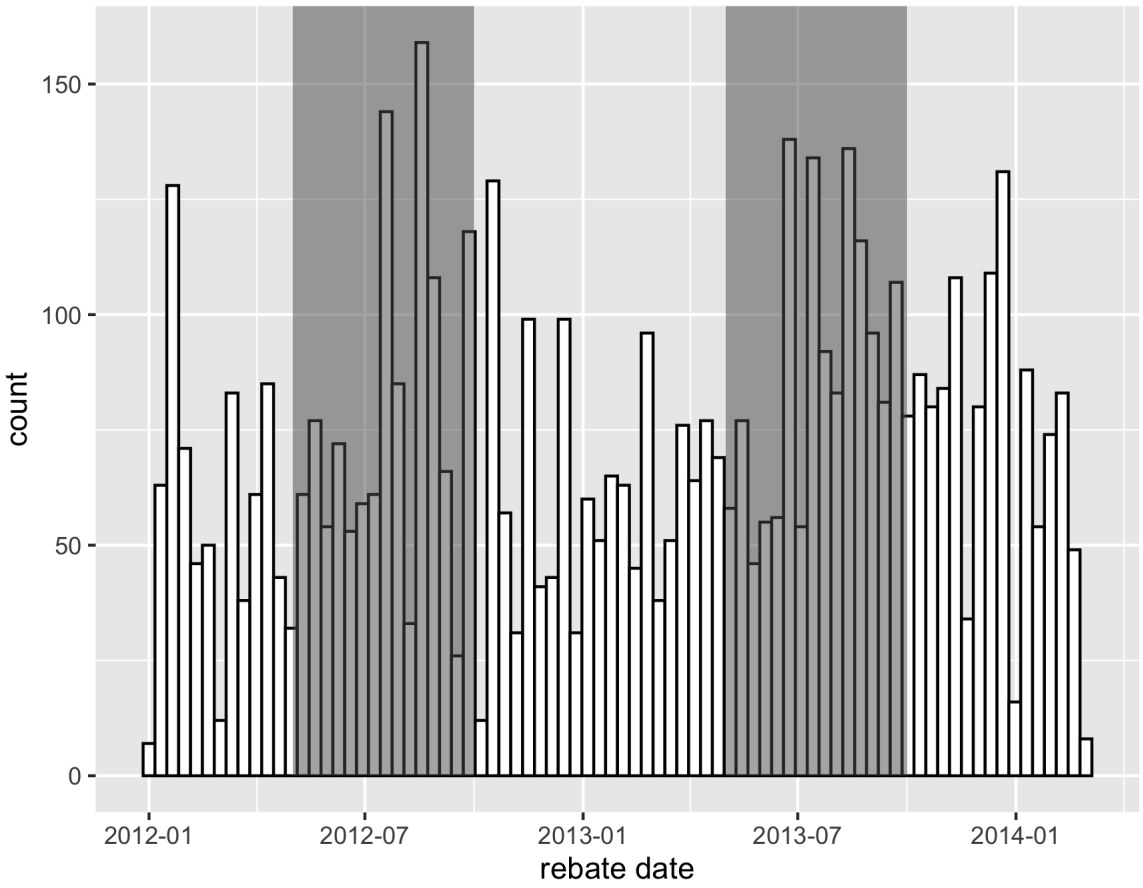
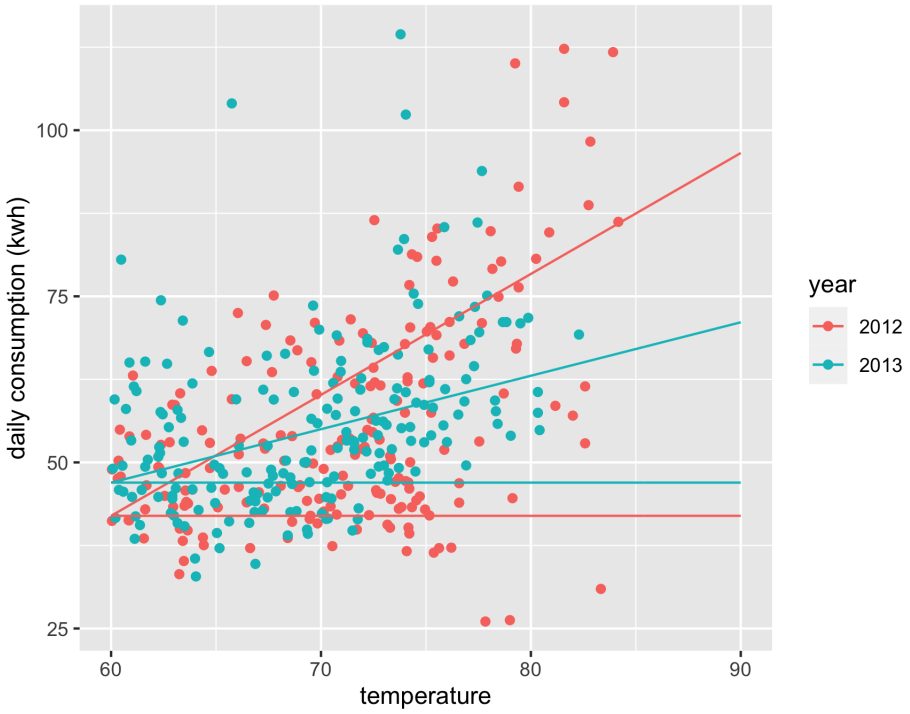
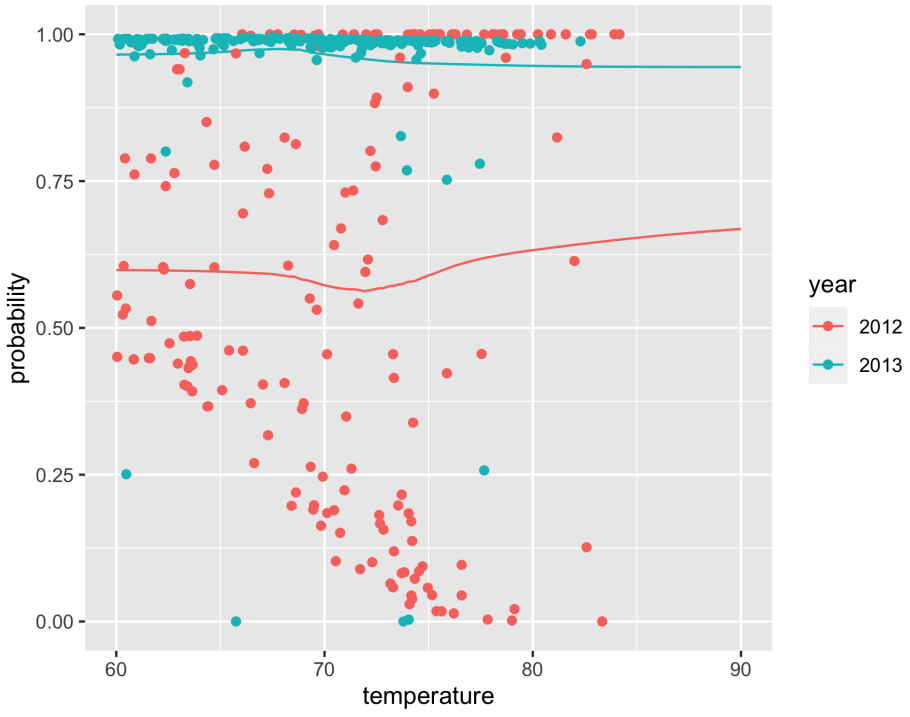


Figure 2.2: Rebate mailing dates histogram

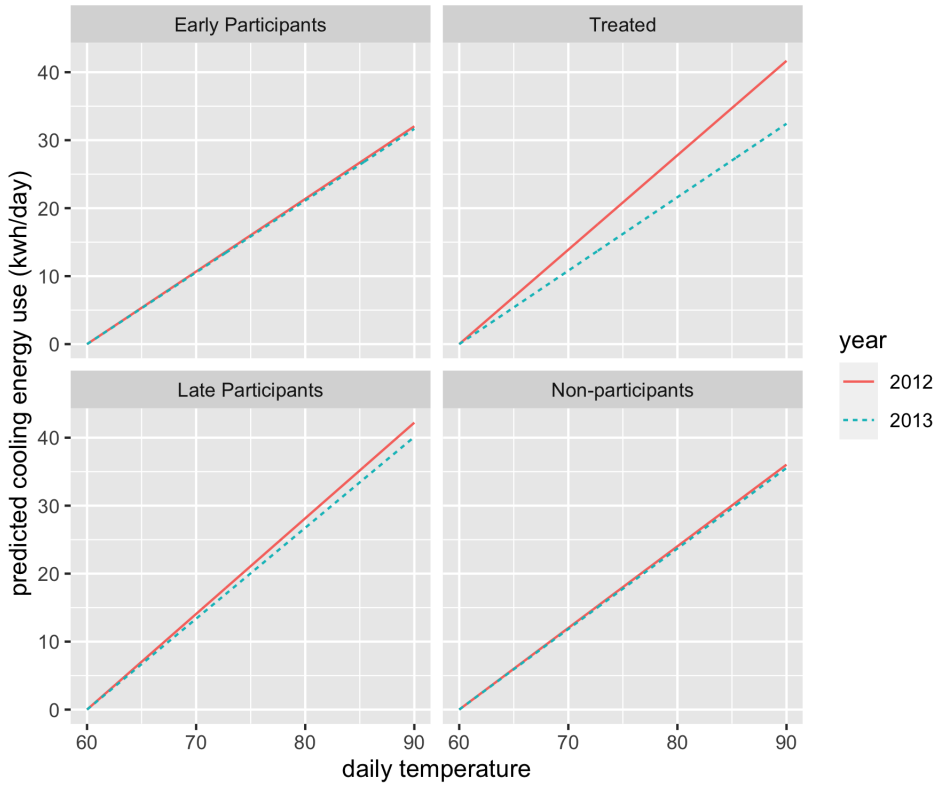


(a) cooling function



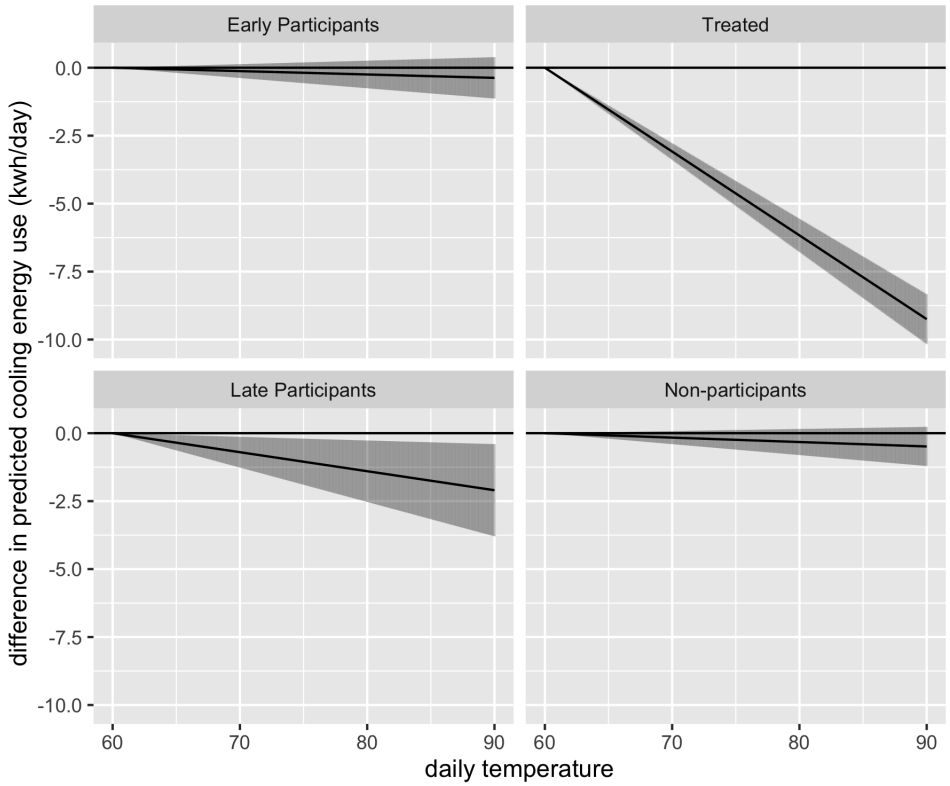
(b) cooling behavior

Figure 2.3: An example fit of the household-level model



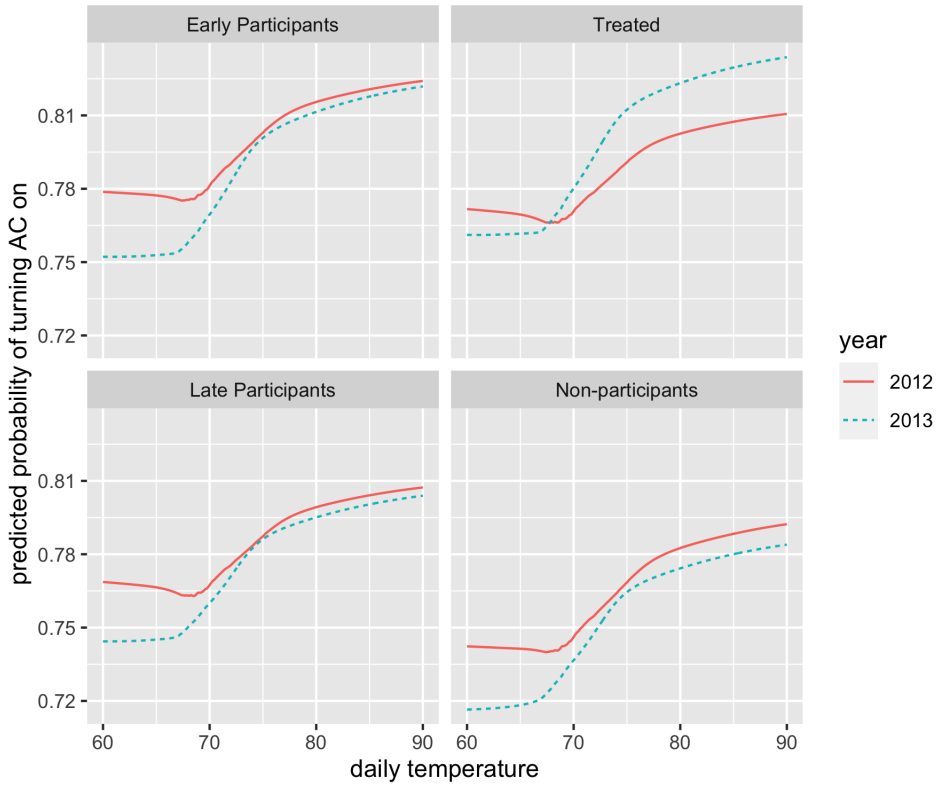
Note: For each premise, we estimate its cooling functions in 2012 and 2013. These linear functions are then aggregated by household types and averages are plotted.

Figure 2.4: Average cooling functions in 2012 and 2013



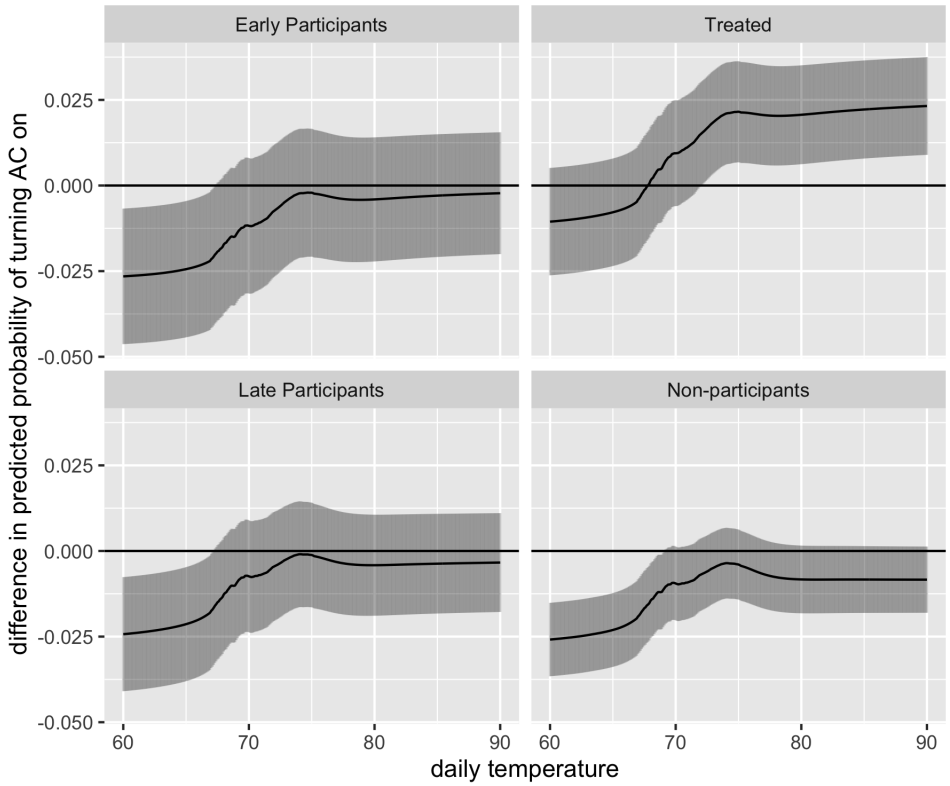
Note: For each premise, we estimate its cooling functions in 2012 and 2013, and then calculate the difference by subtracting the 2012 cooling function from the 2013 cooling function. Differences across households are aggregated by household types and averages are plotted. The shaded areas are 95% confidence intervals generated from two-sample t tests.

Figure 2.5: Difference between 2012 and 2013 cooling lines and two-sample t test C.I.



Note: For each premise, we estimate its behavior functions in 2012 and 2013. These behavior functions are then aggregated by household types and averages are plotted.

Figure 2.6: Average cooling behaviors in 2012 and 2013



Note: For each premise, we estimate its behavior functions in 2012 and 2013, and then calculate the difference by subtracting the 2012 behavior function from the 2013 behavior function. Differences across households are aggregated by household types and averages are plotted. The shaded areas are 95% confidence intervals generated from two-sample t tests.

Figure 2.7: Difference between 2012 and 2013 cooling behaviors and two-sample t test C.I.

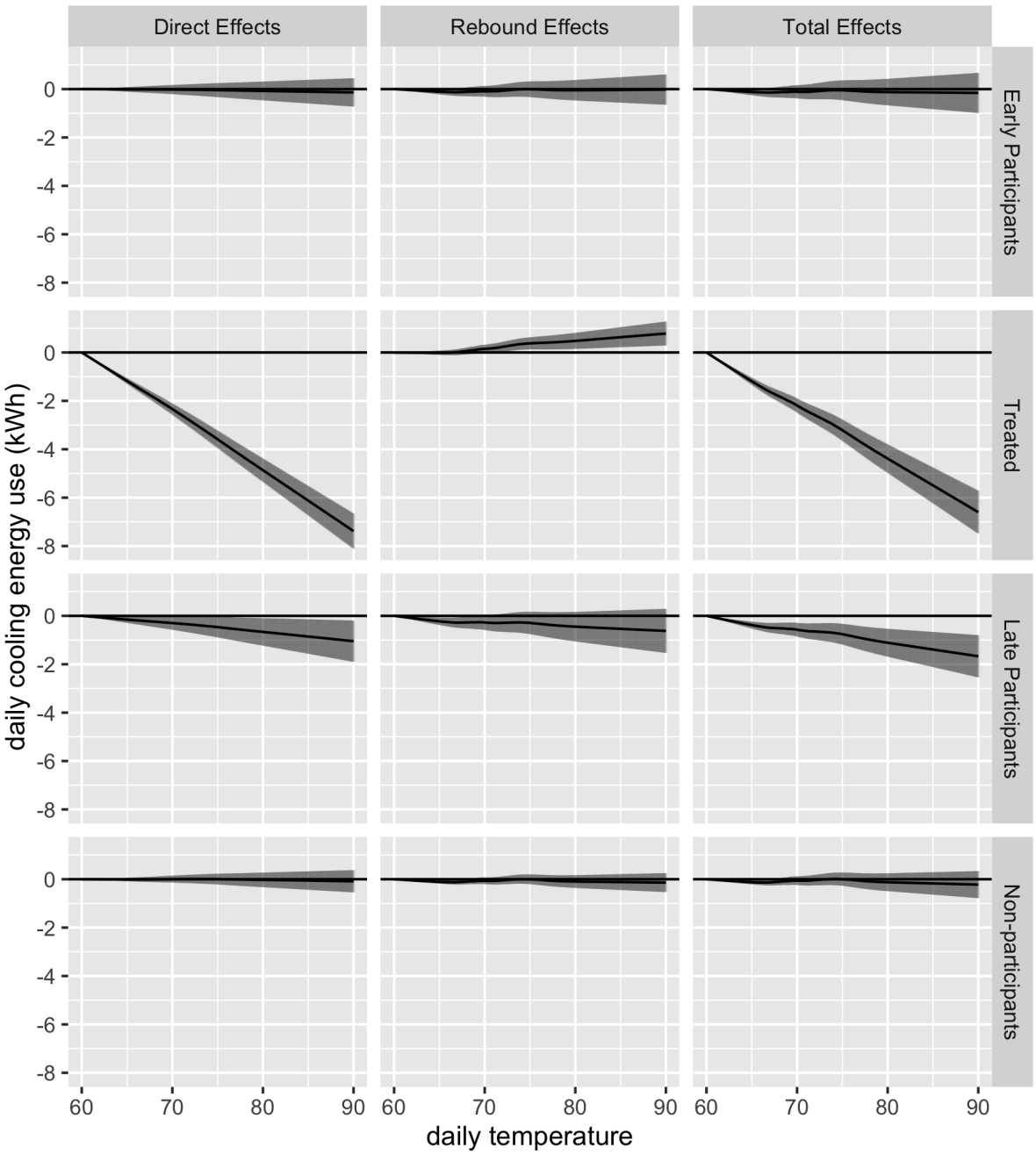


Figure 2.8: Total effects, direct effects and rebound effects by groups (mean and C.I.)

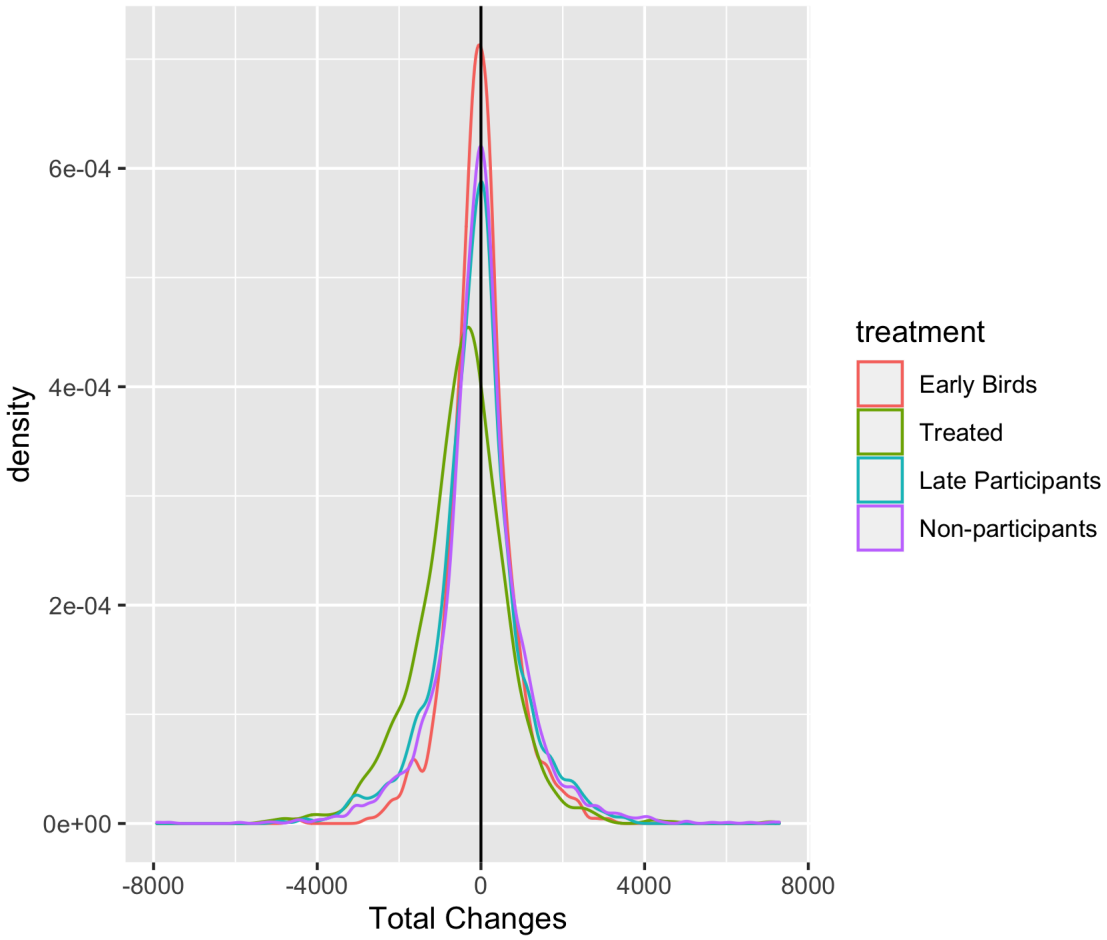
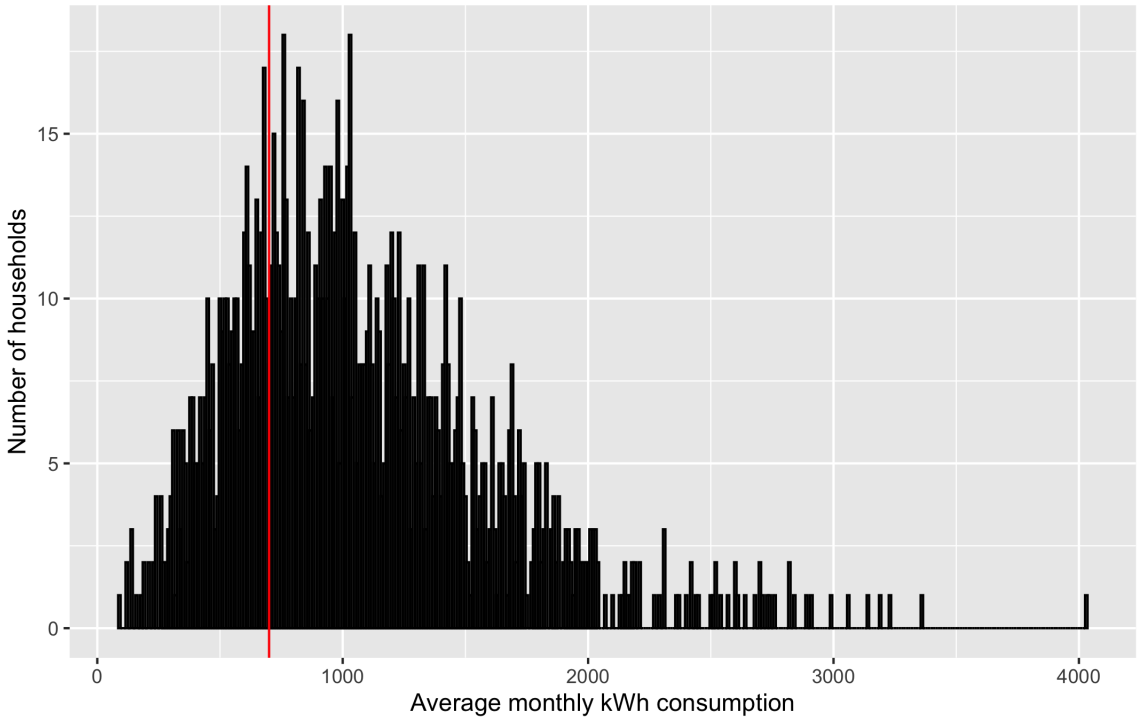


Figure 2.9: Density of total changes prediction by groups



Note: SMUD adopted two rates for its customers during the summer season (June 1 – September 30). For households with monthly consumption below 700 kWh, 9.89 ¢/kWh will be charged, and 18.03 ¢/kWh otherwise. The red line in the plot shows the cutoff kWh, i.e., 700 kWh. Most households consume more than 700 kWh in the 2012 summer season.

Figure 2.10: Number of households by average monthly consumption during June 1st and September 30th, 2012

2.8 Tables

Table 2.1: Mean and standard deviation of household characteristics

Characteristics	Non-participants	Participants
observations	3143	5684
2011 consumption (kWh)	9512.00	10167.46
	(6363.83)	(5016.86)
built in (year)	1976.25	1974.24
	(21.46)	(18.81)
number of bedrooms	3.25	3.36
	(0.97)	(0.75)
number of rooms (total)	6.46	6.61
	(1.70)	(1.37)
number of stories	1.25	1.24
	(0.43)	(0.42)
size (square feet)	1746.69	1794.35
	(678.96)	(642.99)

Source: Sacramento county assessor data

Note: Standard deviations in parentheses

Table 2.2: Mean and standard deviation of household characteristics (participants)

Characteristics	Early Participants	Late Participants	Treated
observations	703	1161	1300
2011 consumption (kWh)	10132.59	10233.51	10118.07
	(5144.41)	(5301.90)	(4899.18)
built in (year)	1972.27	1973.83	1973.21
	(19.15)	(19.52)	(18.85)
number of bedrooms	3.35	3.39	3.35
	(0.78)	(0.76)	(0.74)
number of rooms (total)	6.58	6.69	6.57
	(1.39)	(1.42)	(1.34)
number of stories	1.22	1.25	1.22
	(0.41)	(0.43)	(0.41)
size (sqft)	1776.04	1842.50	1774.73
	(645.57)	(679.28)	(620.61)

Notes: Standard deviations in parentheses

Table 2.3: Number of premises by groups

Group	Obs.
Early Participants	703
Late Participants	1161
Non-participants	3143
Summer 2012	1189
Summer 2013	1331
Treated	1300
Total	8827

Table 2.4: DiD regression results with different comparison groups

	(1)	(2)	(3)	(4)
Total effects	-347.10 ***	-383.62 ***	-442.95 ***	-420.86 ***
	(43.85)	(37.02)	(36.28)	(32.91)
Direct effects	-437.21 ***	-456.99 ***	-497.02 ***	-482.12 ***
	(39.02)	(32.06)	(29.70)	(27.74)
Rebound effects	90.11 **	73.38 **	54.07 **	61.26 **
	(36.27)	(30.50)	(25.97)	(26.20)
Observations	2461	3164	4442	6306
Comparison groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: DiD regression results with different comparison groups (exclude households with negative predicted cooling energy use)

	(1)	(2)	(3)	(4)
Total effects	-364.71 ***	-401.97 ***	-460.95 ***	-438.46 ***
	(43.84)	(36.90)	(35.28)	(32.19)
Direct effects	-449.87 ***	-470.87 ***	-514.80 ***	-498.05 ***
	(38.30)	(31.35)	(27.86)	(26.40)
Rebound effects	85.16 **	68.90 **	53.85 **	59.59 **
	(36.78)	(30.86)	(26.53)	(26.67)
Observations	2401	3095	4231	6058
Comparison groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: DiD regression results with different mixing probability functions

	(1)	(2)	(3)	(4)
constant				
Total effects	-362.17 ***	-386.40 ***	-443.78 ***	-422.42 ***
	(45.55)	(38.57)	(37.48)	(34.16)
Direct effects	-411.30 ***	-433.16 ***	-479.13 ***	-462.01 ***
	(34.18)	(28.35)	(29.02)	(25.95)
Rebound effects	49.13	46.76	35.34	39.59
	(33.85)	(28.93)	(26.74)	(25.61)
linear				
Total effects	-373.11 ***	-427.03 ***	-519.00 ***	-484.76 ***
	(40.07)	(33.59)	(34.02)	(30.33)
Direct effects	-458.32 ***	-489.42 ***	-552.24 ***	-528.85 ***
	(35.99)	(29.92)	(31.13)	(27.53)
Rebound effects	85.21 ***	62.39 ***	33.24 **	44.10 ***
	(15.48)	(13.42)	(16.08)	(14.24)
local weighted average (main results)				
Total effects	-347.10 ***	-383.62 ***	-442.95 ***	-420.86 ***
	(43.85)	(37.02)	(36.28)	(32.91)
Direct effects	-437.21 ***	-456.99 ***	-497.02 ***	-482.12 ***
	(39.02)	(32.06)	(29.70)	(27.74)
Rebound effects	90.11 **	73.38 **	54.07 **	61.26 **
	(36.27)	(30.50)	(25.97)	(26.20)
Observations	2461	3164	4442	6306
Comparison groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: DiD regression results with different spans

	(1)	(2)	(3)	(4)
span = 0.6				
Total effects	-319.22 ***	-363.76 ***	-435.98 ***	-409.09 ***
	(42.32)	(35.80)	(35.89)	(32.33)
Direct effects	-439.44 ***	-468.45 ***	-502.32 ***	-489.71 ***
	(38.95)	(32.53)	(29.96)	(28.19)
Rebound effects	120.22 ***	104.69 ***	66.34 ***	80.62 ***
	(34.75)	(29.79)	(24.57)	(25.17)
span = 0.75 (main results)				
Total effects	-347.10 ***	-383.62 ***	-442.95 ***	-420.86 ***
	(43.85)	(37.02)	(36.28)	(32.91)
Direct effects	-437.21 ***	-456.99 ***	-497.02 ***	-482.12 ***
	(39.02)	(32.06)	(29.70)	(27.74)
Rebound effects	90.11 **	73.38 **	54.07 **	61.26 **
	(36.27)	(30.50)	(25.97)	(26.20)
span = 0.9				
Total effects	-350.06 ***	-379.05 ***	-440.02 ***	-417.32 ***
	(45.00)	(37.94)	(37.15)	(33.75)
Direct effects	-433.40 ***	-451.12 ***	-487.85 ***	-474.17 ***
	(38.50)	(31.70)	(29.26)	(27.47)
Rebound effects	83.34 **	72.07 **	47.83 *	56.86 **
	(37.48)	(31.53)	(25.92)	(26.45)
Observations	2461	3164	4442	6306
Comparison groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.A Additional Regression Tables

2.A.1 Main specification

Table 2.A.1: Details of DiD regression results

	(1)	(2)	(3)	(4)
Equation 2.4				
Total cooling energy use change ($\hat{\beta}_2$)	-113.21 ***	-76.70 ***	-17.36	-39.45 ***
	(31.87)	(23.73)	(19.63)	(14.94)
Total effects	-347.10 ***	-383.62 ***	-442.95 ***	-420.86 ***
	(43.85)	(37.02)	(36.28)	(32.91)
Equation 2.5				
Cooling energy use change	-66.49 **	-46.71 **	-6.68	-21.59 *
- due to cooling curves ($\hat{\beta}_2$)	(28.36)	(20.55)	(16.06)	(12.60)
Direct effects	-437.21 ***	-456.99 ***	-497.02 ***	-482.12 ***
	(39.02)	(32.06)	(29.70)	(27.74)
Equation 2.6				
Cooling energy use change	-46.72 *	-29.98	-10.68	-17.86
- due to behavior changes ($\hat{\beta}_2$)	(26.36)	(19.55)	(14.05)	(11.90)
Rebound effects	90.11 **	73.38 **	54.07 **	61.26 **
	(36.27)	(30.50)	(25.97)	(26.20)
Observations	2461	3164	4442	6306
Comparison groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.A.2 Different rebate dates

In our main text, we define the treated group and late participants based on their rebates mailing dates on or after October 1. In the table below, we report results with the cutoff dates changed to October 15 and November 1. Total effects, direct effects, and rebound effects are comparable (within 95% confidence intervals across specifications). Notice that the significance levels of rebound effects decrease as we move the rebate dates to later dates. This could result from sample sizes becoming smaller as we redefine groups.

Table 2.A.2: DiD regression results with rebate dates

	(1)	(2)	(3)	(4)
October 1st				
Total effects	-347.10 ***	-383.62 ***	-442.95 ***	-420.86 ***
	(43.85)	(37.02)	(36.28)	(32.91)
Direct effects	-437.21 ***	-456.99 ***	-497.02 ***	-482.12 ***
	(39.02)	(32.06)	(29.70)	(27.74)
Rebound effects	90.11 **	73.38 **	54.07 **	61.26 **
	(36.27)	(30.50)	(25.97)	(26.20)
Observations	2461	3164	4442	6306
October 15th				
Total effects	-371.69 ***	-404.96 ***	-452.55 ***	-435.66 ***
	(45.25)	(37.46)	(36.36)	(33.07)
Direct effects	-462.36 ***	-475.62 ***	-502.35 ***	-492.87 ***
	(40.89)	(32.84)	(29.79)	(28.02)
Rebound effects	90.67 **	70.67 **	49.80 *	57.21 **
	(37.78)	(31.11)	(26.07)	(26.40)
Observations	2311	3014	4427	6156
November 1st				
Total effects	-413.94 ***	-437.72 ***	-467.19 ***	-457.24 ***
	(47.53)	(38.68)	(37.95)	(34.56)
Direct effects	-494.85 ***	-498.39 ***	-510.29 ***	-506.27 ***
	(43.73)	(34.32)	(31.10)	(29.39)
Rebound effects	80.91 **	60.67 *	43.09	49.03 *
	(40.88)	(32.94)	(27.41)	(27.86)
Observations	2039	2742	4281	5884
Compare groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.A.3 Exclude negative predicted cooling energy use

Table 2.A.3: Details of DiD regression results (excluding negative cooling energy use)

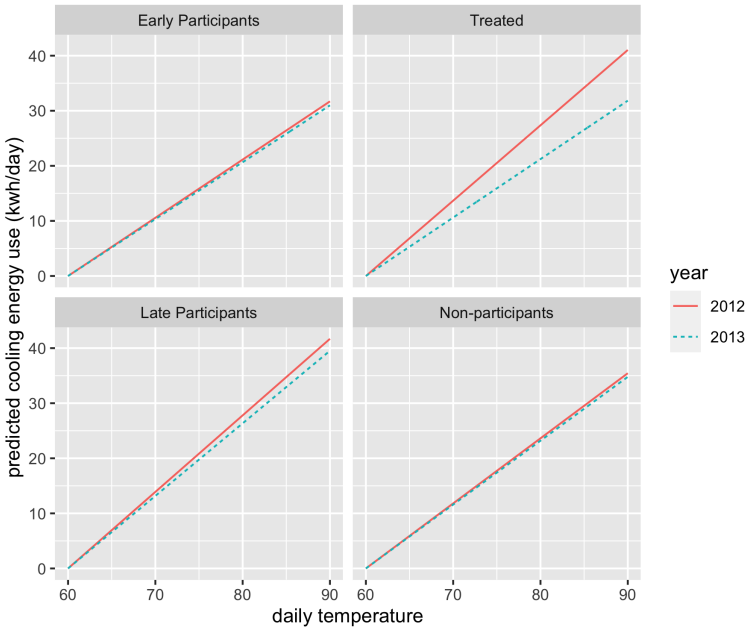
	(1)	(2)	(3)	(4)
Equation 2.4				
Total cooling energy use change ($\hat{\beta}_2$)	-111.42 ***	-74.16 ***	-15.18	-37.68 **
	(31.86)	(23.62)	(19.31)	(14.73)
Total effects	-364.71 ***	-401.97 ***	-460.95 ***	-438.46 ***
	(43.84)	(36.90)	(35.28)	(32.19)
Equation 2.5				
Cooling energy use change	-71.77 ***	-50.77 **	-6.84	-23.60 *
- due to cooling curves ($\hat{\beta}_2$)	(27.84)	(20.07)	(15.25)	(12.08)
Direct effects	-449.87 ***	-470.87 ***	-514.80 ***	-498.05 ***
	(38.30)	(31.35)	(27.86)	(26.40)
Equation 2.6				
Cooling energy use change	-39.65	-23.39	-8.34	-14.08
- due to behavior changes ($\hat{\beta}_2$)	(26.73)	(19.75)	(14.53)	(12.20)
Rebound effects	85.16 **	68.90 **	53.85 **	59.59 **
	(36.78)	(30.86)	(26.53)	(26.67)
Observations	2401	3095	4231	6058
Comparison groups:				
Late Participants	✓	✓		✓
Early Participants		✓		✓
Non-participants			✓	✓

Standard errors in parentheses;

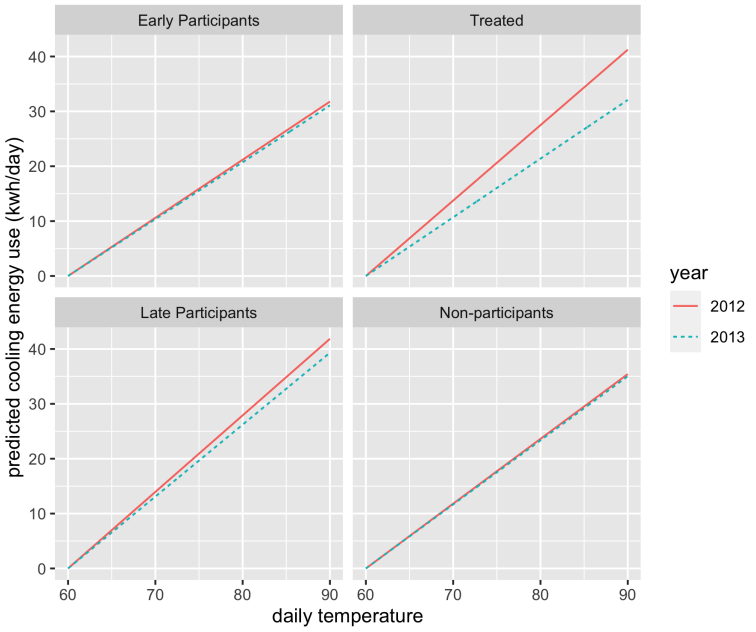
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.B Robustness Check Plots

2.B.1 Cooling behavior functions

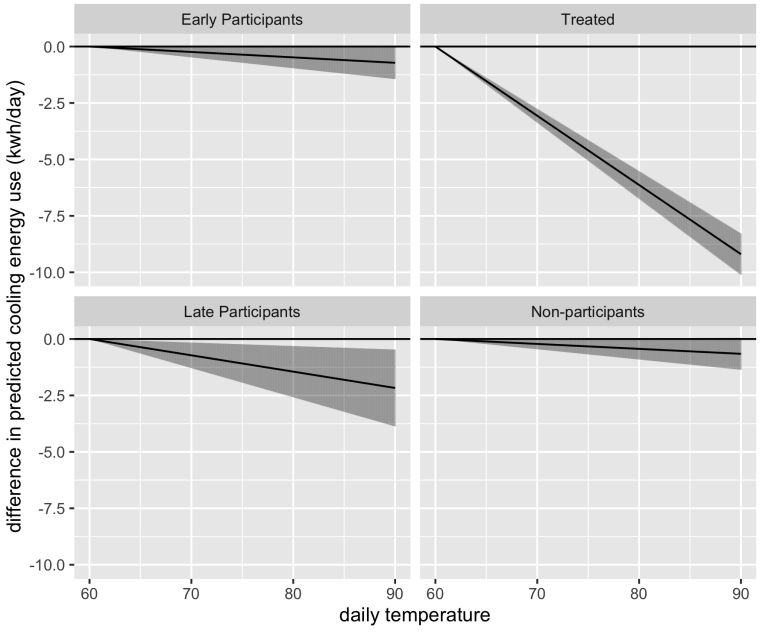


(a) constant

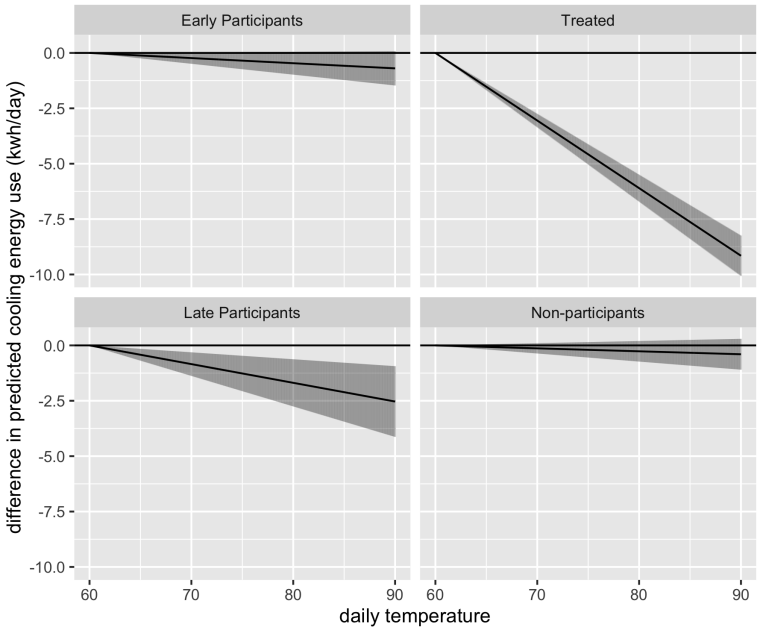


(b) linear

Figure 2.B.1: Cooling functions by groups

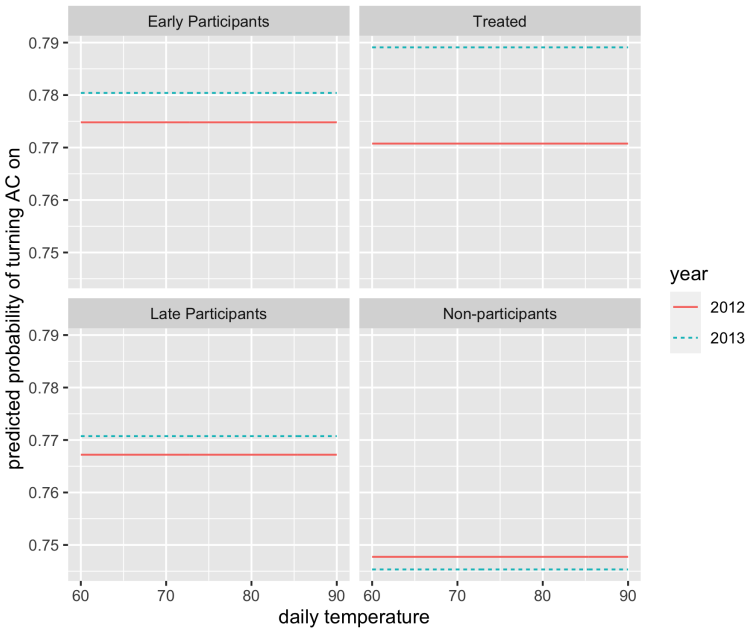


(a) constant

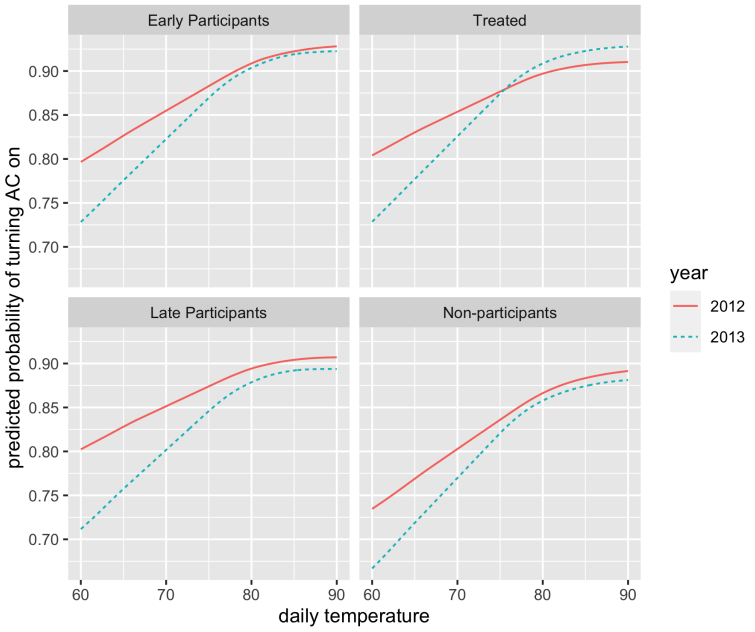


(b) linear

Figure 2.B.2: Differences in cooling functions and C.I.



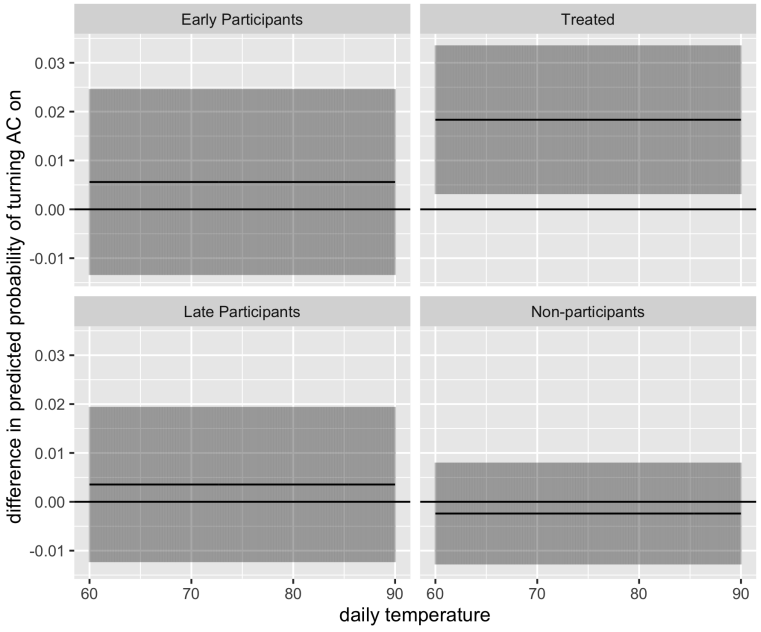
(a) constant



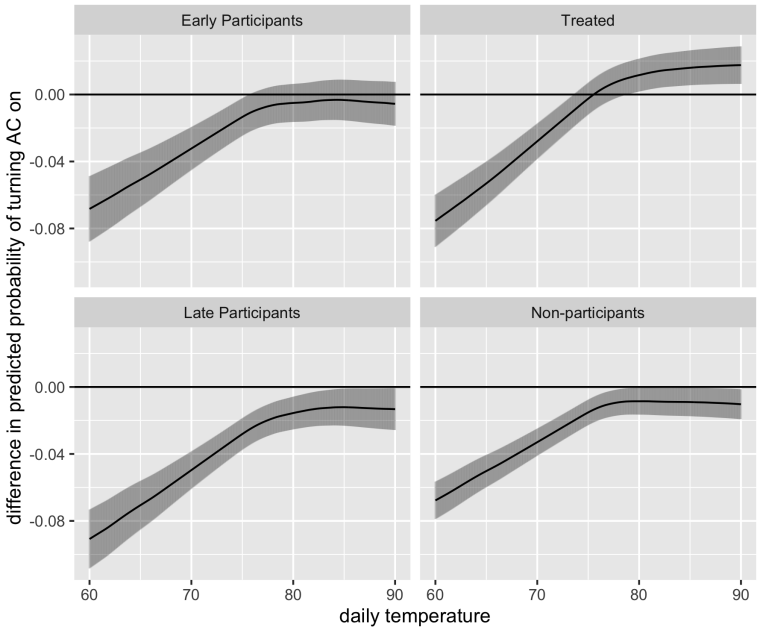
(b) linear [20]

Figure 2.B.3: Cooling behaviors by groups

[20] For each household, we estimate their posterior mixing probability using a linear function in temperatures. Since each household may have their predicted cooling probability greater than 1 at different temperatures, the average of these functions becomes non-linear. Hence, we have concave functions shown above.



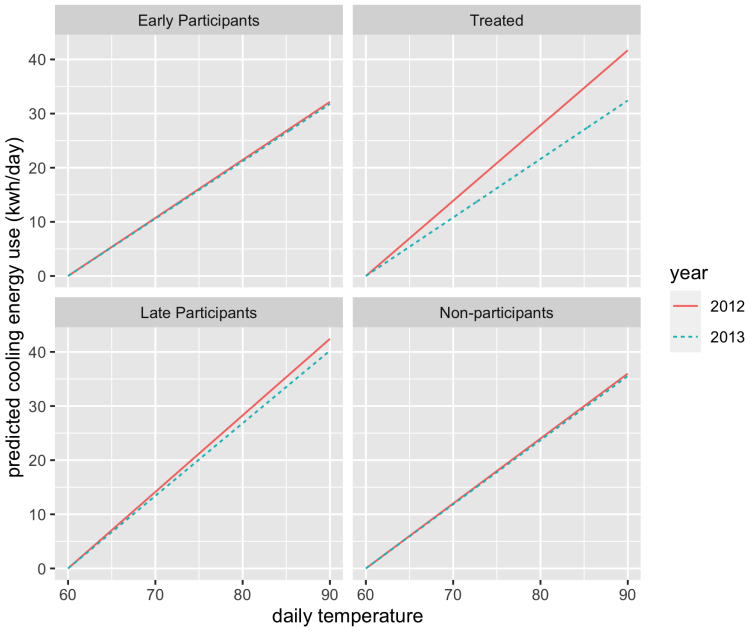
(a) constant



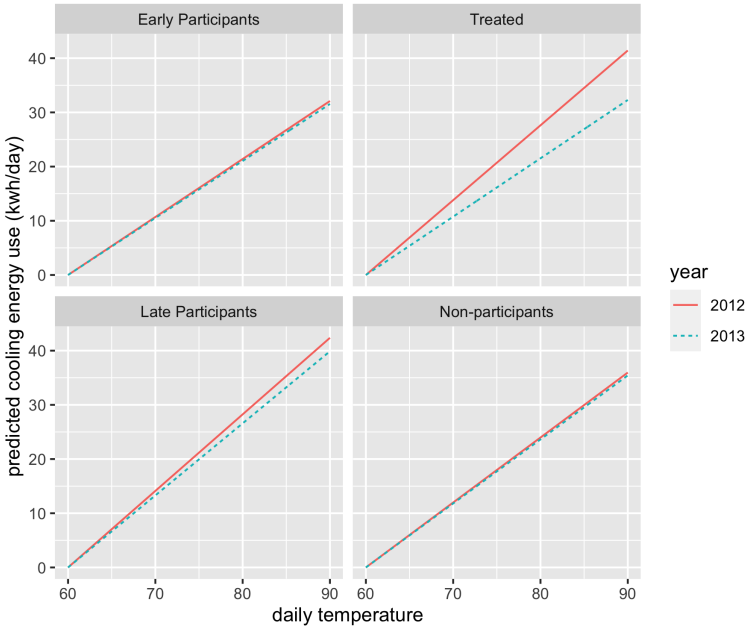
(b) linear

Figure 2.B.4: Differences in cooling behaviors and C.I.

2.B.2 Cooling functions and cooling behaviors for different spans

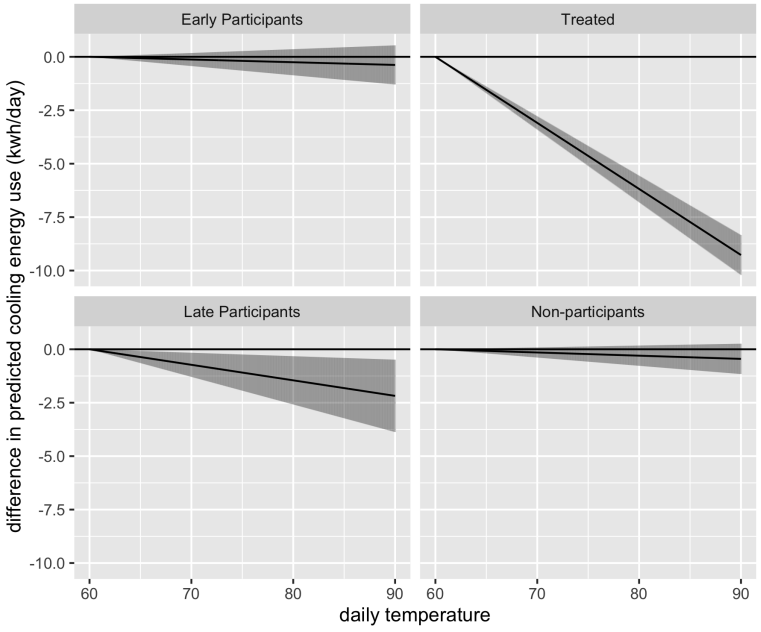


(a) span = 0.6

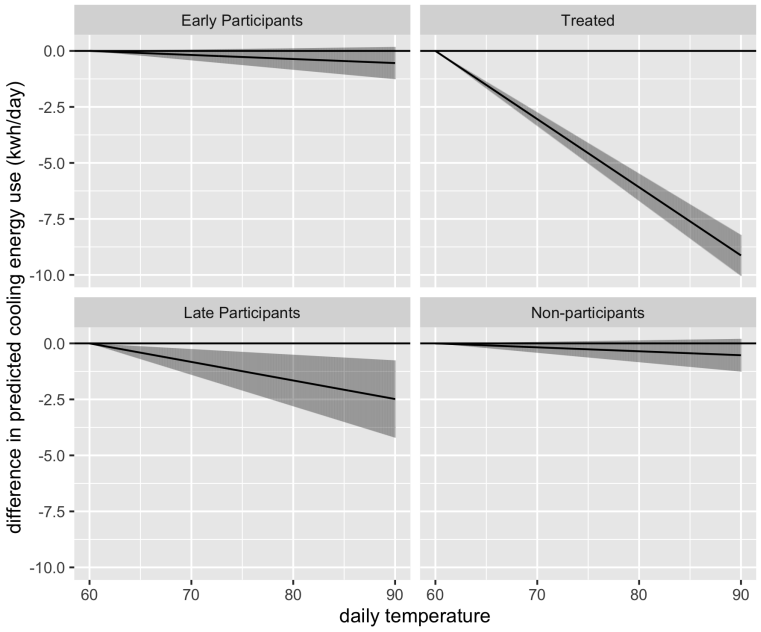


(b) span = 0.9

Figure 2.B.5: Cooling functions by groups

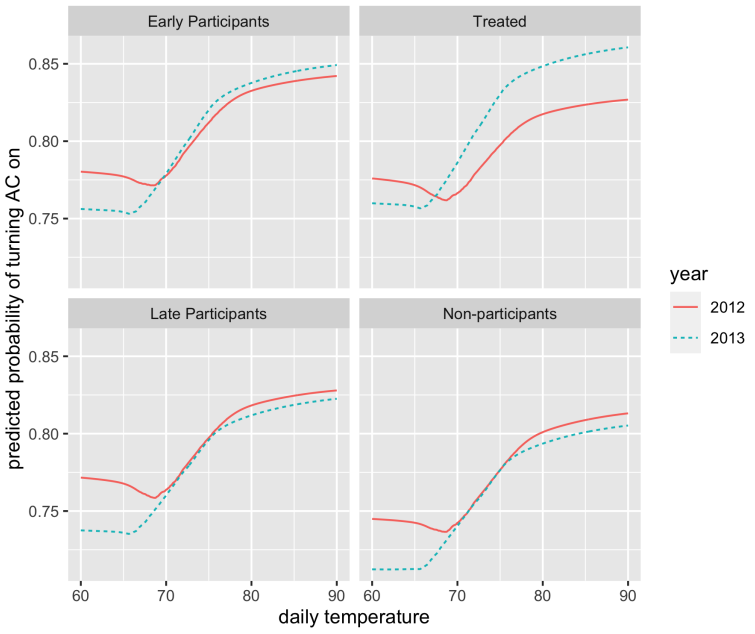


(a) span = 0.6

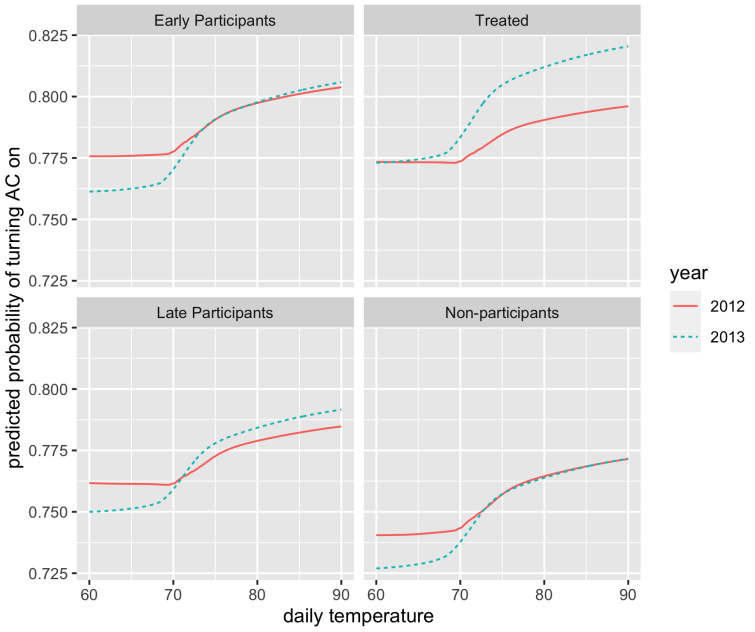


(b) span = 0.9

Figure 2.B.6: Differences in cooling functions and C.I.

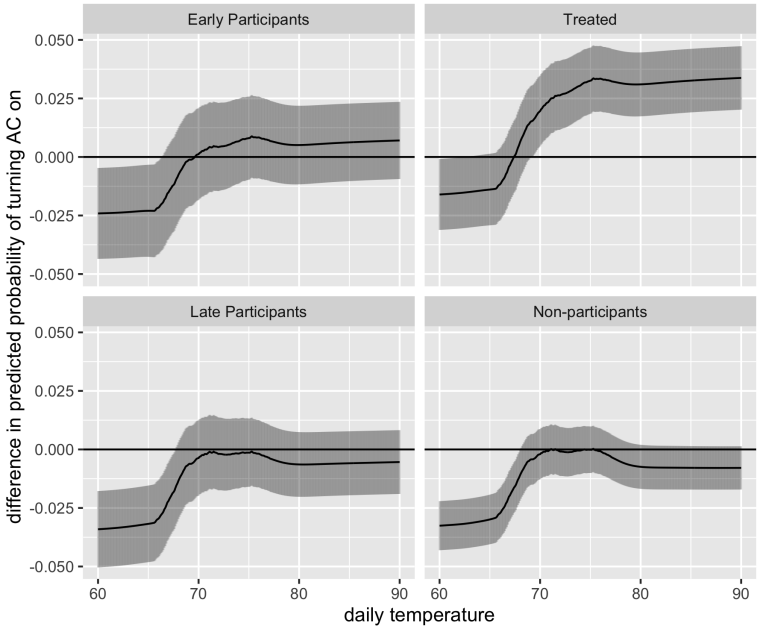


(a) span = 0.6

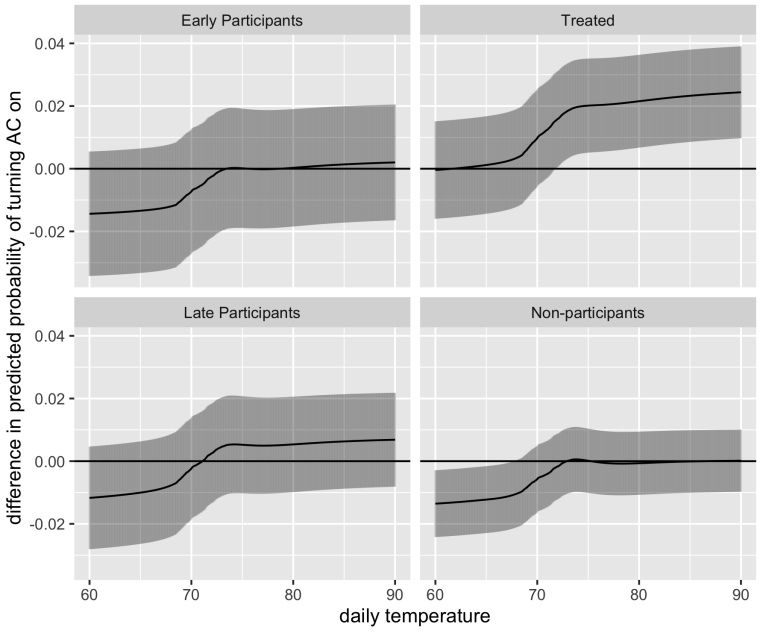


(b) span = 0.9

Figure 2.B.7: Cooling behaviors by groups



(a) span = 0.6



(b) span = 0.9

Figure 2.B.8: Differences in cooling behaviors and C.I.

Essay 3

R&D Lags in Economic Models: Theory and Assessment using Data for U.S. Agriculture

3.1 Introduction

Innovation resulting from organized investments in R&D is at the center of contemporary models of economic growth and is a focus of econometric models of research-induced increases in productivity in agriculture and other industries. Although these branches of applied economics share a common heritage-from work done decades ago by economists like Zvi Griliches, Edwin Mansfield, Jora Minasian, Robert Solow, and Theodore Schultz-, nowadays they employ quite different conceptual and empirical models to represent the process by which today's investments in R&D influence the future time path of productivity and economic growth. These substantial differences in models can be characterized, formally, in terms of differences in the detail of the specification of the R&D lag structure, which transforms measures of past and present investments in R&D into an R&D knowledge stock that affects current productivity.

In a related paper, Alston et al. (2022) flesh out those differences and explore their ori-

gins and implications, taking a broad perspective and drawing on a range of evidence about particular technologies. In the present paper, we focus more narrowly on comparing these alternative models empirically, using a particular data set for U.S. agriculture. These are high-quality data in a comparatively long time-series, which is advantageous for drawing comparisons among the alternative models that differ substantively in terms of their assumptions regarding lag length and shape. Our findings using agricultural data are relevant beyond agriculture; they are informative about comparable relationships for the economy as a whole and the many other industries for which comparably useful data have not been available.

3.2 Data

We compare the alternative models in an application to U.S. agriculture, drawing on long-run data developed specifically for use in models like these by colleagues at the International Science and Technology Practice and Policy (InSTePP) Center at the University of Minnesota. The data used in our analysis include (1) an annual index of U.S. agricultural multifactor productivity (MFP) for the period 1910–2007, obtained from InSTePP; (2) measures of aggregate annual U.S. public agricultural R&D investments and the associated R&D deflator for the period 1890–2007, also sourced from InSTePP; and (3) a purpose-built weather index, which we compute based on crop yield data from the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA), USDA-NASS (2017).

3.2.1 Multifactor Productivity Index

The InSTePP multifactor productivity (MFP) indexes are Fisher ideal discrete approximations of Divisia indexes derived from detailed data on quantities and prices of inputs and outputs in U.S. agriculture. Version 5 of the InSTePP data consists of annual observations of state-specific prices and quantities of 74 categories of outputs and 58 categories of inputs for the 48 contiguous U.S. states from 1949 to 2007, and a corresponding national aggregate. To obtain a longer time series for the national aggregate, MFP is backcast to 1910 using year-to-year changes in the Laspeyres indexes of MFP for the period 1910–1949 from

USDA-ERS (1983). More details on the construction and backcasting of this MFP index can be found in the book by Alston et al. (2010) and the online appendix of Pardey and Alston (2021).

3.2.2 Public Agricultural R&d Investment

InSTePP also provides data on U.S. public agricultural research expenditures for the period 1890–2007, primarily reflecting funding from the federal government to support intramural research undertaken by USDA, and from both federal and state governments to provide for R&D undertaken by the State Agricultural Experiment Stations (SAESs), affiliated with land grant universities.[1] As well as funds from various federal and state government agencies, SAESs obtain funding from industry grants and contracts and income earned from sales, royalties, and various other sources. During the period 1903–1942, USDA intramural research and SAES research contributed almost equally to total public agricultural research spending in the United States. However, since WWII the paths have diverged, and SAES research spending has increasingly exceeded federal intramural research spending, peaking at 75 percent of total public agricultural R&D spending in 2002 (Pardey et al. 2013 and 2017).[2]

3.2.3 Agriculturally Relevant Weather Shocks

Year-to-year fluctuations in crop yields around trend are highly influenced by weather (Beddow et al. 2014), making yield deviations from trend a useful proxy of the transient agricultural productivity effects of weather. Our composite index of crop yield deviations from trend is based on an area-weighted average yield for the years 1940–2007, calculated using yield data for the top 10 crops (by harvested area) taken from USDA-NASS (2017). First, we ranked all 44 field crops in the USDA-NASS (2017) listing according to their average annual harvested areas for the period 1940–2007. Then we selected the top 10 field crops by area (accounting for 78 percent of total harvested area), namely: corn, hay, wheat, soy-

[1]For our analysis in this paper, expenditures were converted to constant (2019-dollar) values using the InSTePP R&D price deflator (unpublished series, updated from Pardey et al. 1989).

[2]More detail on these data and the history of U.S. agricultural R&D investments can be gleaned from Alston et al. (2010, chapter 6) and Pardey et al. (2013).

beans, oats, cotton, sorghum, barley, rice and flaxseed. Since yields vary considerably across crops, we used standardized yields for each crop.[3] These standardized annual crop yields were aggregated by years using as weights each crop's annual share of the total value of production (also from USDA-NASS, 2017). The resulting series was then used in the following time-trend regression:[4]

$$yield_t = \alpha + T_t + T_t^3 + \varepsilon_t \tag{3.1}$$

where $yield_t$ is aggregated standardized yield in year t, and T_t is the time trend created by calendar year minus 1939. We constructed the agricultural weather index in year t as a composite of yield deviations from trend: $yield_t - \widehat{yield_t}$, where: $yield_t$ is the weighted average of the observed yields, aggregated across crops, and $\widehat{yield_t}$ is fitted yields from equation (1). In Figure 3.3 the fitted aggregated yield, $\widehat{yield_t}$ is plotted against the observed aggregated yield, $yield_t$. U.S. agriculture suffered an extended drought in the 1950s (see, e.g., Nace and Pluhowski 1965), and the year 1988 was a severe drought year (see, e.g., GAO 1989), as is apparent in both the yield index and the plot of deviations around it.

3.3 Economic Models of Knowledge Stocks

Economic studies linking R&D to productivity implicitly or explicitly entail a model in which multifactor productivity (MFP_t) depends on flows of services from an R&D knowledge stock, K_t, as well as other factors, X_t:

$$MFP_t = f(K_t; X_t) \tag{3.2}$$

[3]Standardized annual yields were computed by subtracting the mean of the series from each observation and dividing by the standard deviation of the series to reduce the effects of differences in average yields among crops.

[4]Alternative specifications were tried. In particular, we estimated models that included the following terms on the right-hand side of equation (1) besides the constant coefficient α: (a) a linear time trend T_t; (b) a linear time trend T_t and a quadratic time trend ${T_t}^2$; (c) a linear time trend T_t, a quadratic time trend ${T_t}^2$, and a cubic time trend ${T_t}^3$. All these specifications, including the one in equation (1), are not statistically significantly different from one another based on F tests. However, equation (1) results in a slightly higher adjusted R^2 and a slightly lower AIC, which indicates a better fit to our data. Detailed results are included in Appendix Table (3.A.1)

In the typical application, a double-log form is imposed in which the parameters are elasticities:

$$\ln MFP_t = \beta_0 + \beta_K \ln K_t + \beta_X \ln X_t + \varepsilon_t \tag{3.2'}$$

Different assumptions about the processes of creation and utilization of knowledge can be characterized as different parameterizations of the R&D lag structure whereby past and present R&D investments contribute to the stock of knowledge in use today. Applying notation from Alston et al. (2011), the knowledge stock in year t, K_t, can be characterized as:

$$K_t = \sum_{k=0}^{\infty} b_k R_{t-k} \tag{3.3}$$

where b_k is the weight assigned to lag period k, and R_{t-k} is the real (or inflation-adjusted) public agricultural R&D investment in year $t-k$, and (in most cases) these weights sum to one:

$$\sum_{k=0}^{\infty} b_k = 1 \tag{3.4}$$

We are interested in three main categories of models, allowing for some variation within categories, namely: agricultural R&D models, industrial R&D models, and growth theory models. We characterize the differences among these models in terms of differences in the attributes of R&D lag distributions that are imposed implicitly or explicitly: (1) the total lag length, (2) a gestation lag period before research investments begin to contribute to the knowledge stock, (3) restrictions imposed on the functional form of the distribution, and (4) parameters associated with the functional form. In what follows we compare stereotypical examples of the lag structures used in agricultural R&D models, industrial R&D models, and growth theory models both conceptually and in an empirical application using agricultural data.

3.3.1 Agricultural R&D Models

As discussed by Alston et al. (2022a), in applications to U.S. agriculture over the past half century (since Evenson 1967) it has been conventional to model agricultural productivity as

a function of an R&D knowledge stock. The current knowledge stock in use, K_t in year t, is represented by lagged investments in agricultural R&D, with rising and falling lag weights reflecting successive phases of research, development, adoption, depreciation and disadoption of the resulting innovations. Though some have tried free-form weights the great majority of the hundreds of agricultural R&D studies have imposed a structure on the lag distribution so it can be represented by just a few parameters (see, e.g., Alston et al. 2022).[5] As discussed by Pardey et al.(2010), from early beginnings with quite simple models and short lags the models have evolved to allow for longer lags and more complex shapes.

The two predominant models in use nowadays are the 35-year trapezoidal lag distribution model introduced by Huffman and Evenson (1993) thirty years ago, and the 50-year gamma lag distribution model proposed more recently by Alston et al. (2010). Alston et al. (2011) compared these two models applied to U.S. state-level MFP data for the period 1949–2007 from InSTePP, and found in favor of a gamma lag distribution model with a peak lag considerably later than that for the trapezoidal lag model, though otherwise reasonably similar in shape. Both of these models have initial periods of several years with negligible or zero impact of R&D on productivity (a gestation lag or a pre-technology research and development lag) followed successively by a period of rising impact (the adoption lag), and eventually a period of declining impact (reflecting disadoption and depreciation of knowledge in use), truncated to zero at 35 years (the trapezoidal lag distribution model) or 50 years (the 50-year gamma distribution).

In this paper we take the 50-year gamma lag distribution model from Alston et al. (2011) as our starting point—albeit here applied to the national aggregate data rather than state-level data, excluding investments in extension to make for more direct comparability to models applied to other sectors of the economy, and including an additional 10 years of data.[6] Given a 50-year lag, our first knowledge stock observation in 1940 is a weighted

[5] Alston et al. (2022b) report that 540 out of 2,963 estimates of rates of return to agricultural R&D were derived from models using free-form lags.

[6] With a maximum lag of 50 years, and R&D data for the period 1890–2007, we can potentially estimate models for the period 1940–2007. However, Alston et al. (2011) had state-level MFP data beginning in 1949

average of public R&D investments from 1890 to 1940, while the last observation in 2007 is a weighted average of investments from 1957 to 2007. With these measures of knowledge stocks, we can estimate models of MFP using annual data for 1940 to 2007.

Some studies (e.g., Andersen and Song 2013; Khan and Salim 2015) have imposed the specific gamma lag distribution model weights, as estimated by Alston et al. (2011) in other contexts, whether using similar or totally different data. Here, we are using a somewhat different model (i.e., including a different weather index, applied to a single time-series of national aggregate data rather than in a panel of state-level data, and excluding extension expenditures) to model changes in agricultural MFP over a different time period (1940–2007 rather than 1949–2007). Therefore, we opted to re-estimate the gamma lag distribution parameters, using a grid search procedure as done by Alston et al. (2011) across 64 combinations given by 8 values each for the two gamma distribution coefficients. We also tried the Huffman and Evenson (1993) trapezoidal lag model with its specific lag weight structure applied to these different data.

3.3.2 Industrial R&D Models

In models applied to studies of returns to research in other industries, the predominant R&D lag model in use is quite different: it is a perpetual inventory model (see, e.g., Hall 2010; Li and Hall 2018; Serfas et al. 2022). In this model, a proportional declining balance or geometric depreciation rule is used to represent changes in an aggregate stock of knowledge (Griliches 1980, 1986). As described by Alston et al. (2022a), using δ to denote the depreciation rate, and allowing for a gestation lag of g years between research spending and increments to knowledge such that the current gross increment to knowledge is equal to research expenditure g years ago, the aggregate stock of knowledge evolves over time according to:

$$K_t = (1-\delta)K_{t-1} + R_{t-g} = \sum_{s=0}^{\infty}(1-\delta)^s R_{t-s-g} \tag{3.5}$$

Equation (3.5) can be seen as a special case of equation (3.3) in which the entire (infinitely

so they were constrained to estimating models for the period 1949–2007.

long) distribution of lag weights, b_k is represented by one parameter, δ (or two parameters if a nonzero gestation lag is included): $b_{s-g} = (1-\delta)^s$. While it is analytically and empirically convenient, this model imposes strong restrictions on both the length and shape of the R&D-productivity lag relationship.

As typically used, this model allows little or no time for the sequential processes of research, knowledge creation, and the development, diffusion and adoption of technology.[7] The assumed gestation lag is usually very short (if not absent) as is the effective overall lag: in the benchmark case, as described by Li and Hall (2018), $g \leq 2$ (and more often zero) and $\delta = 0.15$.[8] Research has its maximum impact on productivity immediately or almost immediately, and thereafter the lag weights decline rapidly given high assumed rates of knowledge depreciation.

This model seems highly implausible. Why is it so popular? We speculate that the types of firm- or sectoral-level data typically used in models of industrial R&D are not amenable to estimating (and testing among) models with more plausible lag distribution models that have more flexible shapes and longer effective lags. And the perpetual inventory-cum-geometric lag distribution model is quite convenient for applications using data in a very short time-series or a cross-section since the current R&D knowledge stock can be calculated using just the current annual rate of spending, and measures of (or assumptions about) the growth rate of that spending, and the rate of depreciation of the stock.[9]

3.3.3 Growth Models

As described by Jones (1995), the R&D-based models of economic growth associated with Romer (1990), Grossman and Helpman (1991a, b, c), and Aghion and Howitt (1992) all imply scale effects:"... an increase in the level of resources devoted to R&D should increase

[7]This remains so in almost all models of industrial R&D lags, even though some 30 years ago, Griliches (1992, pp. S41–42) declared: "... the more or less contemporaneous timing of such effects is just not possible."

[8]Serfas et al. (2022) compiled 1,464 estimates of rates of return from 128 studies of industrial R&D. Of those 1,464 estimates, 97.3% were based on a perpetual inventory model; 88.2%, did not allow for any gestation lag; 64.4% used a knowledge depreciation rate of $\delta = 15\%$ per year, and another 4.5% used a $\delta > 15\%$ per year.

[9]The knowledge stock in the base period, T, can be approximated as $K_T = R_T/(\delta - \theta)$ where θ is the applicable (often assumed) growth rate of spending on research.

the growth rate of the economy" (Jones 1995, p. 761, emphasis in original). Jones (1995, p. 760) points out that the "... prediction of scale effects is clearly at odds with empirical evidence" and attempts to revise the model to address that deficiency. Others also have found fault with that model and its implausible empirical implications (see, e.g., Jones and Summers 2020).

These issues notwithstanding, the same (unrevised) model from Romer (1990) was employed by Bloom et al. (2020) in recent work that included illustrative applications to several industries, including U.S. agriculture. Specifically, Bloom et al. (2020) presume the current rate of productivity growth is proportional to the current flow of research effort, represented by the number of scientists, measured as research spending divided by an index of the wage rate (which corresponds to R in our notation above). That is, in their equation (1):

$$\frac{\dot{A}_t}{A_t} = \alpha S_t \tag{3.6}$$

In terms of our notation, the growth rate of productivity is measured by MFP, and equation (3.6) can be written as:

$$\frac{\Delta MFP_t}{MFP_t} = \alpha R_t \approx d \ln MFP_t$$

or, equivalently:

$$\ln MFP_t = \alpha R_t + \ln MFP_{t-1} \tag{3.6'}$$

After repeated substitution for the lagged value of equation (3.6'), this can be rewritten as:

$$\ln MFP_t = \alpha \sum_{n=0}^{\infty} R_{t-n} = \alpha K_t \tag{3.7}$$

where the knowledge stock in year t is equal to the accumulated sum of research spending up to period t.

This model assumes research investments have their maximum impact on productivity immediately (i.e., in the same year), without any gestation lag-like the majority of studies

of industrial R&D but in contrast to almost all the studies of agricultural R&D.[10] Further, it assumes these effects that begin immediately continue undiminished, forever. This is significantly different from both studies of industrial R&D (which imply typically rapidly, geometrically declining lag weights) and typical recent studies of agricultural R&D (which allow for rising and falling lag weights over a 35 to 50-year horizon).[11] Also, equation (3.7) is similar to equation (3.2') except that the knowledge stock enters linearly rather than in logarithmic form.

3.3.4 Synopsis of Models - Nested Structure

We have a total of four models to compare, namely: (1) the 50-year (truncated) gamma distribution model (associated with Alston et al. 2011) with its two parameters to be estimated using a grid search, (2) the 35-year trapezoidal model with its specific parameterization (associated with Huffman and Evenson 1993), (3) the geometric model (associated with Hall et al. 2010 among others) using depreciation rates of $\delta = 0.10$ or 0.15, and (4) the Romer-Bloom model (associated with Romer 1990 and Bloom et al. 2020 among others). For the first three of these models (Table 3.1) we impose in common a two-year gestation lag and we limit the maximum length of the R&D lag to 50 years—as was already imposed by Huffman and Evenson (1993), by truncating at 35 years, and implicit as an approximation in the geometric lag model with $\delta = 0.10$ or 0.15 since $0.90^{50} = 0.005$ and $0.85^{50} = 0.0003$. Further, we divide by the 50-year sum of the weights to obtain normalized weights that sum to 1.0.

[10] We are aware of just one study contemplating economic growth models and industrial R&D models together, and ironically it entails an application to agriculture-in Italy. Specifically, Esposito and Pierani (2003) employ a variant of the perpetual inventory model, with a lag distribution characterized by three parameters: (1) the knowledge depreciation rate, (2) a parameter that defines the length of the "gestation period" (before today's R&D has its maximum impact on future productivity), and (3) a parameter that defines the shape of the lag distribution during the gestation period. This lag distribution model seems less plausible than either the gamma lag distribution model or the trapezoidal distribution model, for most cases, but in practice it might yield similar results.

[11] Jones and Summers (2020) begin with a model in the same spirit as Romer (1990) and Bloom et al. (2020) and examine several reasons why the implied benefit-cost ratio may be too high, including a mis-specified R&D lag model. They say "The above baseline assumes that the payoff from R&D investments occurs immediately. Yet there may be substantive delays in receiving the fruits of R&D investments" (Jones and Summers p. 13). "Aggregating across the different types of research, a middle-of-the-road delay estimate may be 6.5 years ..." (Jones and Summers p. 14). These comments refer to the initial R&D lag and adoption processes, but do not address the issue of depreciation of knowledge in use.

For the Romer-Bloom model we do not impose a gestation lag, we do not truncate the lags at 50 years, and we do not impose the restriction that the lag weights sum to 1 (indeed, they are all equal to 1).

Specifically, for each of the gamma, trapezoidal, and geometric lag distribution models we envision the following linear regression model:

$$\ln(MFP_t) = \beta_0 + \beta_1 \ln(K_t) + \beta_2 W_t + T_t + \varepsilon_t \tag{3.8}$$

where MFP_t, K_t, and W_t are, respectively, multifactor productivity, the knowledge stock, and the agricultural weather index in year t, and T_t is a linear time trend (where 1940 is the starting point with $T_t = 1$). In contrast, for the Romer-Bloom model, the knowledge stock enters additively rather than in logarithms:[12]

$$\ln(MFP_t) = \beta_0 + \beta_1 K_t + \beta_2 W_t + T_t + \varepsilon_t \tag{3.8'}$$

In equation (3.8), the growth rate of productivity is proportional to the growth rate of the knowledge stock, and we can interpret β_1 as the elasticity of productivity with respect to the knowledge stock. However, in equation (3.8'), representing the Romer-Bloom model, the growth rate of productivity is simply proportional to the knowledge stock, and the elasticity of productivity with respect to the knowledge stock is equal to $\beta_1 K_t$.[13]

3.4 Time-Series Properties and Lag Model Selection

Ultimately, we aim to estimate the elasticity of productivity with respect to the knowledge stock and the implied benefit-cost ratio (BCR) for agricultural R&D, to see how those esti-

[12] In our regression analysis we try a variant of this model in which we include the Romer-Bloom R&D stock in logarithms rather than levels, to check the importance of this aspect of the difference between this and the other seven models. We thank Aaron Smith for prompting us to take this diagnostic step.

[13] Since the knowledge stock enters linearly and accumulates additively, the estimate of β_1 in equation (3.8') does not depend on the size of the initial knowledge stock in 1939, or how it is estimated, prior to the first observation of MFP, in 1940. Changes in the initial knowledge stock will be absorbed as changes in the intercept without changing any of the slope coefficients. Indeed, for that reason it would be possible to fit that model using data back to 1910—the first year for which we have data available on both MFP and R (and hence, K).

mates compare among the models that differ in terms of the lag specification, and to make an informed choice from among those alternatives. Drawing on Andersen and Song (2013), we propose a systematic method for model selection, which begins with an examination of the time-series properties of the knowledge stocks from each of our lag distribution models, (including 64 gamma lag distribution models, as well as the trapezoidal lag distribution model, two geometric lag distribution models, and the Romer-Bloom model), and the relationship with other variables, namely MFP and the agricultural weather index.

Whether we are estimating (3.8) or (3.8'), we are primarily interested in the estimate of the response of MFP to changes in the knowledge stock, represented by β_1. But for the estimate of β_1 to be meaningful, either the sequences of $\ln(MFP_t)$, $\ln(K_t)$ (or K_t for the Romer-Bloom model), and W_t must be stationary or some linear combination of these variables must be stationary. Otherwise, we will get what Granger and Newbold (1974) call spurious regressions resulting in misleading estimates of β_1. To address this aspect, we first test the stationarity of $\ln(MFP_t)$, its first difference, $\Delta \ln(MFP_t)$, and W_t using the GLS-ADF test (a modified version of the augmented Dickey-Fuller test) proposed by Elliott et al. (1996). Elliott et al. (1996) show that the GLS-ADF test has better power than the standard ADF test when a linear time trend is present in the data (in Figure 3.1 we can see a clear trend in $\ln(MFP_t)$.

The test results are summarized in Table 3.2. In the GLS-ADF test the null hypothesis is that the time series is nonstationary. The results indicate that W_t is stationary. Although $\ln(MFP_t)$ is nonstationary, its first difference (i.e., $\Delta ln(MFP_t)$) is stationary, which indicates $\ln(MFP_t)$ is integrated of order one, $I(1)$. Therefore, to avoid running spurious regressions, requires that $\ln(K_t)$ (or K_t for the Romer-Bloom model) also is $I(1)$ and cointegrated with $\ln(MFP_t)$.[14] The stationarity criterion eliminates 46 of the 64 (i.e., 8×8) parameterizations of the gamma lag model included in our grid search.

Our next step is to test whether $\ln(K_t)$ and $\ln(MFP_t)$ are cointegrated. We opted to

[14]For the Romer-Bloom model, we perform all the time-series tests with respect to K_t instead of $\ln(K_t)$.

perform two cointegration tests: the Johansen (1998) test and the Phillips-Perron (1988) test. The main results are summarized in Table 3.3.[15] In brief, only three gamma lag models (Models 1, 2, and 3 in Table 3.3) pass all the time-series tests. For purposes of comparison, we also include results for another gamma model (Model 4, using parameters from Alston et al. 2011), as well as the trapezoidal lag distribution model (Model 5), the two geometric lag distribution models (Models 6 and 7), the Romer-Bloom model (Model 8), and a logarithmic variant of the Romer-Bloom model (Model 9), none of which has entirely satisfactory time-series properties.

In Table 3.3, columns (4) and (5) refer to the results from applying the same time-series stationarity tests as in Table 3.2, but here with respect to the knowledge stock in order to determine its order of integration. The numbers in columns (4) and (5) indicate we reject the null hypothesis at the specific percentage significance levels shown (i.e., 1%, 5%, or 10%). As discussed above, we require $\ln(K_t)$ to be $I(1)$, which implies we should fail to reject the hypothesis in column (4) but reject the nonstationary hypothesis in column (5). Models 1, 2, 3, 6, and 7 satisfy this criterion for $I(1)$ knowledge stocks.

Next, for the cointegration test, we regress $\ln(MFP_t)$ on $\ln(K_t)$ (or K_t for Model 8) and run Phillips-Perron tests on the residuals. The null hypothesis is that a unit root is present in the residuals. The results are shown in column (6) of Table 3.3. Models 1–5 pass this test but Models 6–8 fail. Finally, in Table 3.3, column (7) we denote that a model passes the Johansen test if it both (1) rejects the hypothesis that there is no cointegrating equation defined by linearly combining $\ln(MFP_t)$ and $\ln(K_t)$ (or K_t for Model 8), and (2) does not reject the hypothesis there is no more than one cointegrating equation defined by the two variables. In other words, $\ln(MFP_t)$ and $\ln(K_t)$ (or K_t for Model 8) form only a single stationary time series. All of the models except the trapezoidal lag model (Model 5) pass the Johansen test. Only the four gamma lag distribution models (Models 1–4) and the logarithmic variant of the Romer-Bloom model (Model 9) pass both the Phillips-Perron and

[15] Further details regarding test statistics, optimal lags, and critical values are included in Appendix Tables (3.A.2)–(3.A.4).

Johansen tests.[16]

Based on this battery of statistical tests, and our strong priors regarding the general structure of the R&D lag, our preferred model is a gamma model with $\gamma = 0.75$ and $\lambda = 0.80$, designated as Model 1 in Table 3.3 and henceforth. In Figure 3.3, we depict the distribution of lag weights assigned to past investments for this model and Models 2 through 7 (i.e., all the models except for the Romer-Bloom model, Model 8). We calculate the peak and average lag for each model and summarize the information in Table **??**. Our preferred gamma model has its peak at year 13, which implies R&D investments make their greatest contribution to the useful knowledge stock 13 years later. Although the lag distribution from this model has a potential lag length of 50 years, its shape is much more similar to that of the trapezoidal model (Model 5, with an imposed lag length of 35 years) than that of Model 4 (with its much longer effective lag length), which was preferred by Alston et al. (2010, 2011).

3.5 Correction for Autocorrelation and Heteroskedasticity

As noted by Anderson and Song (2013) in a similar context, ordinary least-squares (OLS) can provide consistent estimators given stationary relationships among the variables in our specification. However, the estimators and inferences may be biased if the residuals are not independent and identically distributed. The residuals from OLS estimates of equation (8) for the three gamma lag distribution models that pass the time-series tests (Models 1–3) are plotted in Figure 3.5. Although the knowledge stock differs across these models, the plots of the residuals are similar. From 1940 to 1970, the models seem to suffer from autocorrelation, and each of the three residual plots exhibits a wide apparent range of variance. Accordingly, we conduct tests for heteroskedasticity and autocorrelation, and ultimately utilize estimates from regression models with corrections for heteroskedasticity and autocorrelation.

Table 3.5 summarizes the results from formal tests for heteroskedasticity and autocorre-

[16]These cointegration tests are strictly relevant (only for the models that had satisfied the stationarity tests).

lation for our preferred gamma lag distribution model (Model 1). Test results for the other models can be found in Appendix Tables (3.A.5) and (3.A.6). To test for heteroskedasticity we use the White test (for nonlinear forms of heteroskedasticity) and the Breusch-Pagan test (for linear forms of heteroskedasticity). The null hypothesis is that the errors have a constant variance. Since we do not have a large data set, we implemented Wooldridge's (2015) version of the White test to save degrees of freedom. The results indicate that we might have a nonlinear heteroskedasticity problem in the error terms: we reject the null hypothesis of constant error variance null under the White test at a 5% significance level, though not at 1%. In our OLS and DOLS regressions, we use Newey-West heteroskedasticity and autocorrelation consistent (HAC) standard errors.[17] In the regressions that use the Cochrane-Orcutt procedure or the Prais-Winsten procedure to correct for autocorrelation we use Eicker-Huber-White standard errors to correct for heteroskedasticity. These corrections will not affect the point estimates of β_1. However, they will affect the confidence intervals.

Durbin-Watson (DW) and Breusch-Godfrey (BG) tests are considered to test for autocorrelation in the error terms. The DW test can be used to test for a first-order autoregressive structure in models where the error terms follow a normal distribution and the regressors are strictly exogeneous. The BG test can be used to test for higher orders of autoregressive structures, and it also does not require regressors to be strictly exogenous. From the test results, we reject the null hypothesis that there is no autocorrelation in the error terms up to the specified lags. The evidence strongly suggests that we should correct for at least first-order autocorrelation in the error terms, and we consider three options for doing so: dynamic OLS (Stock and Watson, 1993), the Cochrane-Orcutt procedure, and the Prais-Winsten procedure. The dynamic OLS method does not specify the order of autocorrelation while the latter two procedures take care of AR(1) serial correlation in the errors. Compared with the Cochrane-Orcutt procedure, the Prais-Winsten procedure has the advantage of preserving the first observation in the data transformation step and, given a small sample size, it might

[17] We use Newey-West HAC estimators with pre-whitening and a finite sample adjustment. See the R manual on the function NeweyWest for details.

produce different results.

3.6 Regression Results, Elasticities and Benefit-Cost Ratios

The dynamic OLS estimates are preferred because this procedure corrects for more general autocorrelation problems. In Table 3.6, we report complete results for all nine lag distribution models estimated using dynamic OLS with Newey-West HAC standard errors.[18] Corresponding results estimated with OLS, the Cochrane-Orcutt and the Prais-Winsten procedures are reported in Appendix Tables (3.A.7) – (3.A.9). In Table 3.6, the preferred model is Model 1, and the other models are presented for purposes of comparison and to illustrate the consequences of model specification choices.

Estimation methods procedures might also matter for findings. In Table 3.7, we focus on the estimates of the elasticities of MFP with respect to the knowledge stock from those same regressions across the nine lag distribution models and the four different estimation procedures.

3.6.1 Elasticity Estimates

In the OLS estimates (Table 3.7, column (1)), all of the coefficients except one are estimated quite precisely with small standard errors, they are all in keeping with prior expectations and the relevant economic theory, and they are quite similar across all but one of the eight models. The notable and sole exception is the coefficient on the knowledge stock in the Romer-Bloom model (Model 8) for which the point estimate in column (3) is not statistically significantly

[18] Stock and Watson (1993) did not provide an empirical procedure for selecting the optimal lags and leads for the first difference of the cointegrated regressors (i.e., $\Delta \ln(K_t) = \ln(K_t) - \ln(K_{t-1})$ for equation (3.8) and $\Delta K_t = K_t - K_{t-1}$ for equation (3.8')). We follow a data-driven procedure as used by Choi and Kurozumi (2012) to select the optimal lags and leads. In particular, we first define the maximum numbers of lags and leads using $floor(4 \times (T/100)^{1/4})$, where $floor(x)$ is a floor function which gives the greatest integer less than or equal to x, and T is the total number of years in our data. The resulting maximum number of lags and leads in our sample is three. Next, we run dynamic OLS regressions with different combinations of lags and leads ($\Delta \ln(K_{t\pm i}) = \ln(K_{t\pm i}) - \ln(K_{t\pm i-1})$), where $i \in \{1, 2, 3\}$ and compute the BIC for each model. The model with the optimal lags and leads will produce the smallest BIC.

different from zero at a 5% level of significance.[19]

Comparing the estimates across columns (1)–(4) of Table 3.7, we see that the corrections for autocorrelation had mostly modest effects on the point estimates of the elasticities productivity with respect to the knowledge stock. The notable exceptions are meaningful increases, especially with dynamic OLS, in the point estimates of the elasticities for Models 3, 4, and 9-models with larger mean lags compared with the other methods. However, even when they did not increase the point estimates, the corrections for autocorrelation affected the standard errors on some of the estimates of elasticities of productivity with respect to the knowledge stock, sufficiently to change the inferences in some cases—notably in Models 3 and 4.

When the Cochrane-Orcutt procedure is employed (column (2)) nothing changes very much compared with OLS (column (1)), but more pronounced differences are observed when the Prais-Winsten procedure is employed (column (3)), reflecting the combination of smaller estimated standard errors and larger point estimates of elasticities. Now, compared with the OLS estimates (column (1)), the elasticity of productivity with respect to the knowledge stock in Model 2 is statistically significantly different from zero at the 1% level, rather than 5%; the elasticities in Models 3, 4, 6, 7 and 9 are statistically significantly different from zero at the 5% level, but not 1%; and the elasticity from Model 8 is now statistically significant at the 10% level but not 5%.

Compared with OLS, the dynamic OLS regressions result in a slightly less-precise and less statistically significant estimate of the elasticity of productivity with respect to the knowledge stock in the preferred model (from 1% to 5%) and significantly more precise estimates of the elasticities from three other models: elasticities from Models 3, 4, and 9 that were not statistically significant are now all significantly different from zero at the 1% level. This might be because the leads and lags of the first differences of the knowledge

[19]Recall, in the Romer-Bloom model, the elasticity of productivity with respect to the knowledge stock is equal to $\beta_1 K_t$, where in the other models the elasticity of productivity with respect to the knowledge stock is equal to β_1.

stock variables absorb short-term noise, resulting in more precise estimators of the long-run cointegration relationships (i.e., elasticities).

In what follows we focus on the estimates obtained using the dynamic OLS regressions. The elasticities reported in column (4) of Table 3.7 for lag distribution Models 1–7 and 9 range from 0.201 to 0.386. This is a remarkably narrow range given the considerable differences in the shapes of the lag distributions across the models. The largest value comes from Model 3. The point estimate of the elasticity for Model 1 is essentially the same across the different estimation methods (0.290 for OLS, 0.306 for Cochrane-Orcutt, 0.307 for Prais-Winsten, and 0.277 for dynamic OLS).

3.6.2 Benefit-Cost Ratios

As first suggested by Griliches (1958) the gross annual benefits from productivity growth are approximately equal to the product of the gross value of production, V, and the growth rate of multifactor productivity, MFP:

$$B_t = \frac{\Delta MFP_t}{MFP_t} V_t \approx d \ln MFP_t V_t \tag{3.9}$$

In equation (3.8), growth in multifactor productivity is linked to research spending through the knowledge stock, K:

$$d \ln MFP_t = \beta_1 d \ln K_t \; where \; K_t = \sum_{k=0}^{\infty} b_k R_{t-k} \tag{3.10}$$

An increase in research spending in the current year, t, by ΔR_t will give rise to a stream of benefits from its effects on the time path of the stock of knowledge and thus productivity:

$$\left.\frac{\Delta MFP_{t+k}}{MFP_{t+k}}\right|_{\Delta R_t} = \beta_1 \left.\frac{\Delta K_{t+k}}{K_{t+k}}\right|_{\Delta R_t} = \beta_1 \frac{b_k \Delta R_t}{K_{t+k}} \tag{3.11}$$

Given a discount rate of 100 r percent per year, the discounted present value of benefits from an increase in research spending in the current year, t, is therefore equal to:

$$PVB_t = \sum_{k=0}^{\infty} \frac{\Delta MFP_{t+k}}{MFP_{t+k}}\bigg|_{\Delta R_t} V_{t+k}\ (1+r)^{-k} = \sum_{k=0}^{\infty} \beta_1\ b_k\ \Delta R_t \frac{V_{t+k}}{K_{t+k}}(1+r)^{-k} \qquad (3.12)$$

Hence, the benefit-cost ratio (BCR) for an increase in research spending in year t by ΔR_t is:[20]

Table 3.8 presents the BCRs and 95% confidence intervals computed using a real discount rate of 3 percent per year (i.e., $r = 0.03$) for the seven models that yielded sensible results (Models 1–7). The BCRs were computed using equation (20) with the elasticities estimated by OLS (column (1)), or with corrections for autocorrelation using either the Cochrane-Orcutt procedure (column (2)) or the Prais-Winsten procedure (column (3)), the last of which is the preferred estimation procedure.

The first row of Table 3.8 refers to results for our preferred lag distribution model (Model 1). In column (4), the dynamic OLS estimate of the BCR is 23.4, and it is statistically significantly different from zero. In columns (1), (2) and (3), the alternative estimation procedures yield very similar estimates (24.5, 25.9, and 26.0) for Model 1. The same is true for the estimates of BCRs for Models 2–7: looking across columns in any specific row the estimates are very similar. Reflecting the results with respect to elasticities, the OLS estimates of BCRs are mostly statistically significantly different from zero. However only four lag distribution models yield statistically significant BCR estimates across all estimation procedures: the preferred gamma model (Model 1), the almost identical trapezoidal model (Model 5), the somewhat similar gamma model (Model 2), and the geometric model with 10% depreciation (Model 6). Recall, of these four models, only Models 1 and 2 satisfy the

[20] Reflecting the difference between equations (3.8) and (3.8'), the benefit-cost ratio for the Romer-Bloom model is different, specifically

$$BCR_t = \frac{PVB_t}{R_t} = \beta_1 \sum_{k=0}^{\infty} V_{t+k}(1+r)^{-k}$$

However, given the negative value for the estimated coefficient, this equation is not applicable to our estimates so we do not report any estimates for that model.

time-series conditions required for robust estimates.

The preferred estimates of BCRs are those in column (4), based on the dynamic OLS regressions. They are all statistically significantly different from zero across eight of the nine lag distribution models, the exception being the Romer-Bloom Model. Looking down column (4), among the eight lag distribution models the estimated BCRs range from 18.5 (Model 7) to 27.3 (Model 3), a surprisingly narrow range at first blush. These differences in BCRs reflect the effects of differences in elasticities combined with different lag shapes and discounting—a lag distribution with a greater mean lag, everything else equal, will have a smaller BCR and more so if the discount rate is greater.

Compared with Model 1 (our preferred gamma lag model, with a BCR of 23.4), Model 5 (the trapezoidal lag model) has a a slightly larger BCR (25.2) reflecting its combination of a slightly larger elasticity and a somewhat shorter lag—it peaks at years 9 to 15 compared with year 13 for Model 1. In contrast, Model 2 has a smaller elasticity and a somewhat longer lag resulting in a somewhat smaller BCR (20.5). The other two gamma lag distribution models (Models 3 and 4) both have substantially longer lags. In spite of its relatively long lag, Model 3 has the highest BCR (27.3) reflecting its considerably larger elasticity, while Model 4 has both a smaller elasticity and a long lag and a relatively small BCR (18.9). Finally, while they too have smaller elasticities the two geometric lag distribution models (Models 6 and 7) also have much shorter lags, with offsetting effects on the estimated BCRs (21.0 and 18.5 respectively).

The results in Table 3.8 were obtained with a discount rate of 3 percent per year, which we think is appropriate for this application. In Table 3.9 we show the consequences of alternative discount rates applied to compute BCRs with the dynamic OLS estimates of the elasticities. In every row of this table, as we increase the discount rate from a very low ($r = 0.001$, 0.1 percent per year) to a very high ($r = 0.10$, 10 percent per year) the estimated BCR falls—for our preferred model it falls from a high of 36.3 to a low of 9.8, still quite impressive, bracketing the BCR in column (2) of 23.4. But this effect is more pronounced for

the models with the longer lags, with implications for the relative sizes of the BCRs across models and even the ranking. In column (4), with a 10 percent discount rate the geometric models (Models 6, and 7) now have BCRs greater than that for the preferred model (Model 1).

One of the striking features of these results is the strong similarity and substantial size and statistical significance of the estimated BCRs regardless of whether the underlying lag distribution model is fully consistent with priors (Models 1 through 5) or totally at odds with them (Models 6, 7 and 9). That this is so can be partly understood by considering the extensive discussion of "Plausibility of Estimates" in the book by Alston et al. (2010, pp. 423–435). As they show there, the annual value of agricultural productivity growth is many times greater than annual public spending on agricultural R&D. Hence, if the productivity growth is attributed entirely to that R&D spending, the BCR must be very large even if a long R&D lag is imposed. This aspect of the problem is common across all the models and the variants tried.[21]

3.7 Conclusion

The work in this paper was inspired by our observation of striking differences in the stereotypical R&D lag distribution models used by economists studying the economics of agricultural R&D, compared with economists studying the economics of R&D in other industries or modeling economic growth more broadly. Specifically, applications to agricultural R&D typically employ a 35- to 50-year R&D lag distribution model, with phases of rising and falling lag weights as innovations are progressively created, introduced, adopted and eventually replaced. In contrast, stereotypical models of industrial R&D and popular economic growth models entail much less likely assumptions of very short or nonexistent R&D lags and very

[21] A related consideration discussed by Alston et al. (2010) is the potential for attribution bias resulting from the omission of potentially relevant explanatory variables such as agricultural extension knowledge stocks (as included in the model used by Alston et al. 2010, 2011), private agricultural R&D knowledge stocks (as tried but without any empirical success by Huffman and Evenson 2006) or other sources of technology spillovers such as international agricultural R&D or other U.S. industrial R&D. These omissions might have resulted in upward-biased estimates of the elasticities and, consequently, the BCRs from all the models. However, we suspect these biases would be modest, for the reasons given by Alston et al. (2010).

high (at one extreme) or zero (at the other extreme) rates of knowledge depreciation.

We set out to codify these differences into a nested structure and conduct a comparative assessment of their empirical consequences using a comparatively rich data set in relatively long time-series. Our data set for U.S. agriculture is similar to those used by others in several recent studies (see, e.g., Alston et al. 2011; Andersen and Song 2013; Baldos et al. 2019), and it is a context in which we have strong priors, based on detailed evidence of various forms, about the credibility of models that entail assumptions of very short or nonexistent R&D lags and extreme assumptions about the rate of knowledge depreciation (see, e.g., Pardey et al., 2010 and Alston et al. 2010, 2011, and 2022).

The quantitative results are surprising in some ways. First, apart from the Romer-Bloom model, which implied a negative effect of R&D on productivity, the other seven models all yielded rather similar estimates of elasticities of productivity with respect to the R&D knowledge stock and, in turn, quite comparable estimates of BCRs—all well within the range of widely accepted status quo estimates (see, e.g., the review by Fuglie 2018). If someone had naïvely estimated just one (any one) of these models by OLS, viewing the estimates uncritically they might have been well pleased by the seemingly strong and apparently credible results.

But even if they work well as statistical models, two of these models (the geometric lag distribution models, Models 6 and 7) are not at all plausible in the application to U.S. agriculture, if anywhere (Alston et al 2022a). Further, four of the seven models (Models 4, 5, 6, and 7) fail to satisfy time-series (stationarity and cointegration) tests. Notably, we rejected (Model 4) the specific gamma lag distribution model that was found to be best in the similar application by Alston et al. (2010, 2011). Fortunately, we were able to estimate two models (Models 1 and 2) that performed well as statistical models, that were not inconsistent with our prior expectations regarding the likely length and shape of the R&D lag distribution, and that yielded plausible and statistically significant results within the range of reasonable expectations.

Interestingly, our preferred gamma lag distribution model is quite different in its general shape from the model preferred by Alston et al. (2010, 2011). Even though it allows for a longer, 50-year, lag it has a very similar overall shape to the shorter (35-year) Huffman and Evenson (1993) trapezoidal lag model. It also appears to be very similar in shape to the preferred gamma lag model identified by Baldos et al. (2019), which also implies a similar value for the BCR. Moreover, the estimate of the elasticity of productivity with respect to the knowledge stock (0.28) from our preferred model is remarkably close to what Baldos et al. (2019) estimated (0.29) using a Bayesian hierarchical approach.[22]

Most researchers are not in a position to estimate a flexible lag distribution model and test among alternatives in the ways we have done here using data for U.S. agricultural R&D. Instead, almost all studies linking R&D to productivity simply impose untested assumptions about the length and shape of the R&D lag, which can potentially have profound implications for the results. Some such assumptions are inevitable and indeed desirable. Forty years ago, Zvi Griliches (1979, p. 106, emphasis in original) suggested "... it is probably best to assume a functional form for the lag distribution on the basis of prior knowledge and general considerations and not to expect the data to answer such fine questions."

But Griliches does not tell us what to assume about the form for the R&D lag distribution, and at least some groups of economists—in particular, those measuring returns to industrial R&D or using R&D—based models of economic growth—have made a habit of imposing assumptions in their lag distribution models that seem to be significantly at odds with reality. It should be possible to make better judgments about this aspect of the model specifications. Getting these ideas right matters. Even though they might seem superficially similar—in terms of the estimated elasticities and BCRs—the alternative lag distribution models can have profoundly different implications for our economic understanding of the linkages between investments in R&D, productivity, and economic growth, and the temporal structure of those linkages.

[22]Fuglie (2018, p. 437) reports an elasticity of MFP with respect to national public agricultural R&D equal to 0.30 for North America, computed as the average of estimates across seven studies.

References

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): 323–351.

Alston, Julian M, and Philip G Pardey. 2021. "The economics of agricultural innovation." *Handbook of agricultural economics*, 5: 3895–3980.

Alston, Julian M, Matthew A Andersen, Jennifer S James, and Philip G Pardey. 2010. *Persistence Pays: US Agricultural Productivity Growth and the Benefits from Public R & D Spending.* Springer: New York, NY.

Alston, Julian M, Matthew A Andersen, Jennifer S James, and Philip G Pardey. 2011. "The economic returns to US public agricultural research." *American Journal of Agricultural Economics*, 93(5): 1257–1277.

Alston, Julian M, Philip G Pardey, and Xudong Rao. 2022b. "Payoffs to a half century of CGIAR research." *American Journal of Agricultural Economics*, 104(2): 502–529.

Alston, Julian M, Philip G Pardey, Devin Serfas, and Shanchao Wang. 2022a. "Research Lags Redux: Reconciling Industrial and Agricultural R&D Lag Models." InSTePP Working Paper. St Paul, MN: International Science and Technology Practice and Policy Center, University of Minnesota, in process.

Andersen, Matthew A, and Wenxing Song. 2013. "The Economic impact of public agricultural research and development in the United States." *Agricultural Economics*, 44(3): 287–295.

Baldos, Uris Lantz C, Frederi G Viens, Thomas W Hertel, and Keith O Fuglie. 2019. "R&D spending, knowledge capital, and agricultural productivity growth: A Bayesian approach." *American Journal of Agricultural Economics*, 101(1): 291–310.

Beddow, Jason M, Philip G Pardey, and Terrance M Hurley. 2014. "Reassessing the Effects of Weather on Agricultural Productivity." Selected paper presented at the Agricultural and Applied Economics Association's 2014 AAEA Annual meeting, Minneapolis, MN.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb. 2020. "Are ideas getting harder to find?" *American Economic Review*, 110(4): 1104–44.

Choi, In, and Eiji Kurozumi. 2012. "Model selection criteria for the leads-and-lags cointegrating regression." *Journal of Econometrics*, 169(2): 224–238.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. "Efficient Tests for an Autoregressive Unit Root." *Econometrica*, 64(4): 813–836.

Esposti, Roberto, and Pierpaolo Pierani. 2003. "Building the knowledge stock: Lags, depreciation, and uncertainty in R&D investment and link with productivity growth." *Journal of Productivity Analysis*, 19(1): 33–58.

Evenson, Robert. 1967. "The contribution of agricultural research to production." *Journal of Farm Economics*, 49(5): 1415–1425.

Fuglie, Keith. 2018. "R&D capital, R&D spillovers, and productivity growth in world agriculture." *Applied Economic Perspectives and Policy*, 40(3): 421–444.

Granger, Clive WJ, and Paul Newbold. 1974. "Spurious regressions in econometrics." *Journal of econometrics*, 2(2): 111–120.

Griliches, Zvi. 1958. "Research costs and social returns: Hybrid corn and related innovations." *Journal of political economy*, 66(5): 419–431.

Griliches, Zvi. 1964. "Research expenditures, education, and the aggregate agricultural production function." *The American Economic Review*, 54(6): 961–974.

Griliches, Zvi. 1979. "Issues in assessing the contribution of research and development to productivity growth." *The bell journal of economics*, 92–116.

Griliches, Zvi. 1980. "Returns to research and development expenditures in the private sector." In *New developments in productivity measurement*. 419–462. University of Chicago press.

Griliches, Zvi. 1986. "Productivity, R and D, and Basic Research at the Firm Level in the 1970's." *The American Economic Review*, 76(1): 141–154.

Griliches, Zvi. 1992. "The search for R&D spillovers." *National Bureau of Economic Research Working Paper Series*, , (w3768).

Griliches, Zvi. 1996. *R&D and productivity: The econometric evidence.* University of Chicago Press.

Grossman, Gene M, and Elhanan Helpman. 1991a. *Innovation and growth in the global economy.* MIT press.

Grossman, Gene M, and Elhanan Helpman. 1991b. "Quality ladders and product cycles." *The Quarterly Journal of Economics*, 106(2): 557–586.

Grossman, Gene M, and Elhanan Helpman. 1991c. "Quality ladders in the theory of growth." *The review of economic studies*, 58(1): 43–61.

HALL, Bronwyn H. 2005. "Measuring the Returns to RD: The Depreciation Problem." *Annales d'Ãconomie et de Statistique*, , (79/80): 341–381.

Hall, Bronwyn H, Jacques Mairesse, and Pierre Mohnen. 2010. "Measuring the Returns to R&D." In *Handbook of the Economics of Innovation*. Vol. 2, 1033–1082. Elsevier.

Huffman, Wallace E, and Robert E Evenson. 1993. "Science for agriculture: a long term perspective." Iowa State University, Department of Economics.

Huffman, Wallace E, and Robert E Evenson. 2006. "Do formula or competitive grant funds have greater impacts on state agricultural productivity?" *American Journal of Agricultural Economics*, 88(4): 783–798.

InSTePP. 2020. *Unpublished InSTePP R&D Deflators.* InSTePP (International Science and Technology Practice and Policy) center, St Paul, MN: University of Minnesota.

Johansen, Søren. 1988. "Statistical analysis of cointegration vectors." *Journal of economic dynamics and control*, 12(2-3): 231–254.

Jones, Benjamin F, and Lawrence H Summers. 2020. "A Calculation of the Social Returns to Innovation." National Bureau of Economic Research Working Paper 27863.

Jones, Benjamin F, Lawrence H Summers, et al. 2020. *A calculation of the social returns to innovation.* Vol. 27863, National Bureau of Economic Research.

Jones, Charles I. 1995. "R&D Based Models of Economic Growth." *Journal of Political Economy*, 103(4): 759–784.

Jones, Charles I. 2005. "Growth and Ideas." In . Vol. 1 of *Handbook of Economic Growth*, , ed. Philippe Aghion and Steven N. Durlauf, 1063–1111. Elsevier.

Khan, Farid, and Ruhul Salim. 2015. "The Public R&D and Productivity Growth in Australian Broadacre Agriculture: A Cointegration and Causality Approach." Australian Agricultural and Resource Economics Society 2015 Conference (59th), February 10-13, 2015, Rotorua, New Zealand 204432.

Li, Wendy CY, and Bronwyn H Hall. 2020. "Depreciation of business R&D capital." *Review of Income and Wealth*, 66(1): 161–180.

Lobell, David B, Uris Lantz C Baldos, and Thomas W Hertel. 2013. "Climate adaptation as mitigation: the case of agricultural investments." *Environmental Research Letters*, 8(1): 015012.

Nace, Raymond Lee, and Edward J Pluhowski. 1965. *Drought of the 1950's with Special Reference to the Mid-continent.* US Government Printing Office.

Pardey, Philip G., and Julian M. Alston. 2021. "Unpacking the Agricultural Black Box: The Rise and Fall of American Farm Productivity Growth." *The Journal of Economic History*, 81(1): 114–155.

Pardey, Philip G., and Vincent Smith. 2017. *Waste Not, Want Not—Transactional Politics, Research and Development and the US Farm Bill.* Policy Brief in the series "Agricultural Policy in Disarray, Reforming the Farm Bill." Washington D.C.: American Enterprise Institute.

Pardey, Philip G., Barbara Craig, and Michelle L. Hallaway. 1989. "U.S. agricultural research deflators: 1890-1985." *Research Policy*, 18(5): 289–296.

Pardey, Philip G., Julian M. Alston, and Connie Chan-Knag. 2013. *Public Food and Agricultural Research in the United States: The Rise and Decline of Public Investments, and Policies for Renewal.* AGree Report. Washington D.C.: Agree.

Pardey, Philip G., Julian M. Alston, and Vernon W. Ruttan. 2010. "Chapter 22 - The Economics of Innovation and Technical Change in Agriculture." In *Handbook of the Economics of Innovation, Volume 2.* Vol. 2 of *Handbook of the Economics of Innovation*, , ed. Bronwyn H. Hall and Nathan Rosenberg, 939–984. North-Holland.

Pardey, Philip G., Matthew A. Anderson, and Julian M. Craig, Barbara J. Alston. 2009. *Primary Data Documentation U.S. Agricultural Input, Output, and Productivity Series, 1949–2002 (Version 4).* InSTePP Data Documentation. St Paul, MN: International Science and Technology Practice and Policy Center, University of Minnesota.

Pardey, Philip G., Matthew A. Anderson, and Julian M. Craig, Barbara J.and Alston. 2014. "InSTePP United States Production Accounts, version 5." St Paul,

MN: International Science and Technology Practice and Policy, 2014. Available from www.instepp.umn.edu/united-statesVersion%205.

Phillips, Peter C. B., and Pierre Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika*, 75(2): 335–346.

Plastina, Alejandro, and Lilyan Fulginiti. 2012. "Rates of return to public agricultural research in 48 US states." *Journal of Productivity Analysis*, 37(2): 95–113.

Rao, Xudong, Terrance M Hurley, and Philip G Pardey. 2019. "Are agricultural R&D returns declining and development dependent?" *World Development*, 122: 27–37.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, 98(5): S71–S102.

Serfas, Devin, Julian M. Alston, and Philip G. Pardey. 2022. "The Returns to Industrial RD: A Comprehensive Meta-Review and Reassessment of the Evidence." InSTePP Working Paper. St Paul, MN: International Science and Technology Practice and Policy Center, University of Minnesota, in process.

Stock, James H., and Mark W. Watson. 1993. "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems." *Econometrica*, 61(4): 783–820.

Ugur, Mehmet, Eshref Trushin, Edna Solomon, and Francesco Guidi. 2016. "RD and productivity in OECD firms and industries: A hierarchical meta-regression analysis." *Research Policy*, 45(10): 2069–2086.

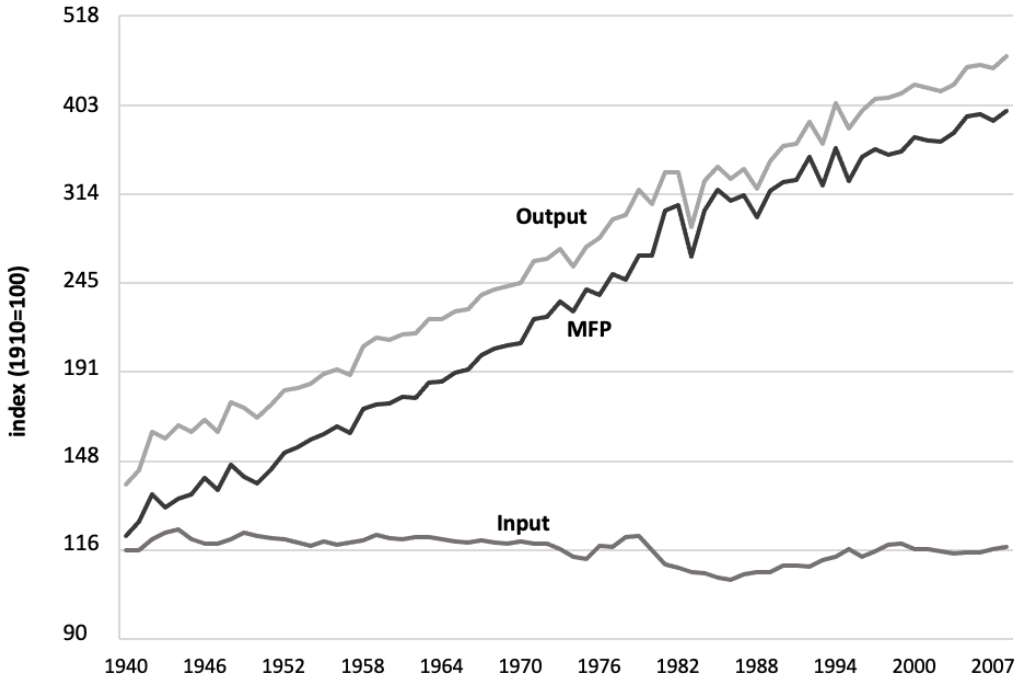
United States General Accounting Office (GAO). 1989. *Crop Production Outlook for Post-drought Recovery.* Briefing Report to the Chairman, Subcommittee on Agricultural Research and General Legislation, Committee on Agriculture, Nutrition, and Forestry, U.S. Senate. GAO/RCED-89-161BR, June.

USDA-ERS (U.S. Department of Agriculture, Economic Research Service). 1981. *Economic Indicators of the Farm Sector: Production and Efficiency Statistics, 1979. Statistical bulletin*, U.S. Department of Agriculture, Economics and Statistics Service.

USDA National Agricultural Statistics Service. 2017. "Nass — Quick Stats." USDA National Agricultural Statistics Service. https://data.nal.usda.gov/dataset/nass-quick-stats. Accessed 2020-12-31.

Wooldridge, Jeffrey M. 2015. *Introductory Econometrics: A Modern Approach.* 6th Ed. Mason, OH: South-Western Cengage Learning.

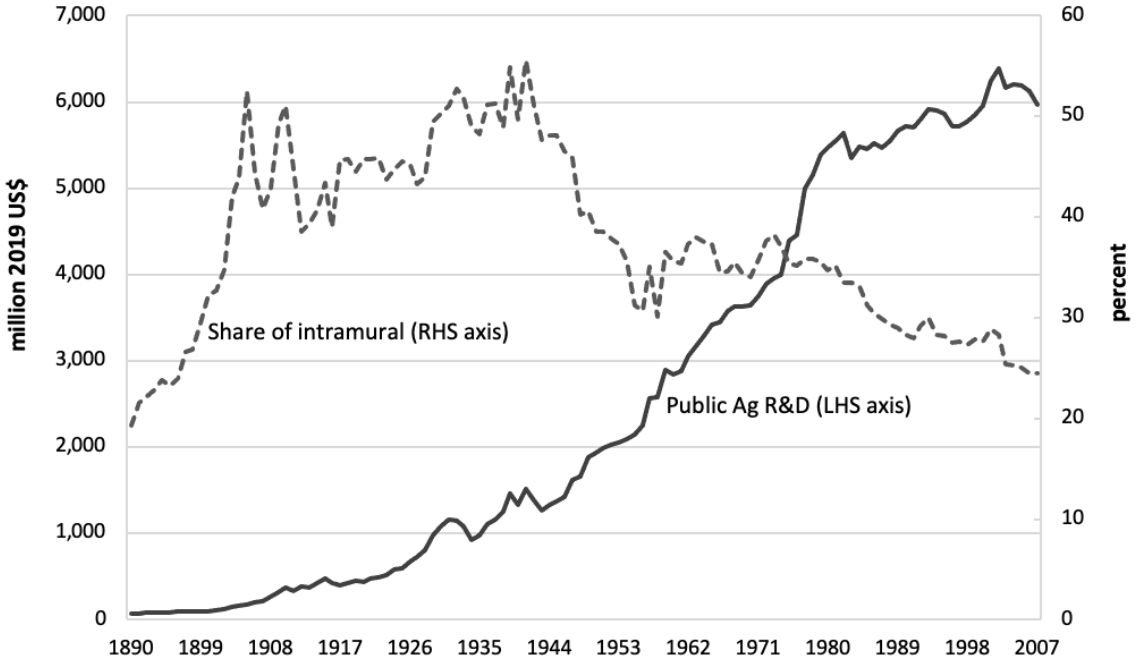
3.8 Figures



Sources: University of Minnesota, InSTePP Center compilation drawing on InSTePP Production Accounts, version 5, augmented with data from USDA-ERS (1983).

Notes: Plots are natural logs of the respective indexes with base year 1910=100. Y axis reports the actual index values.

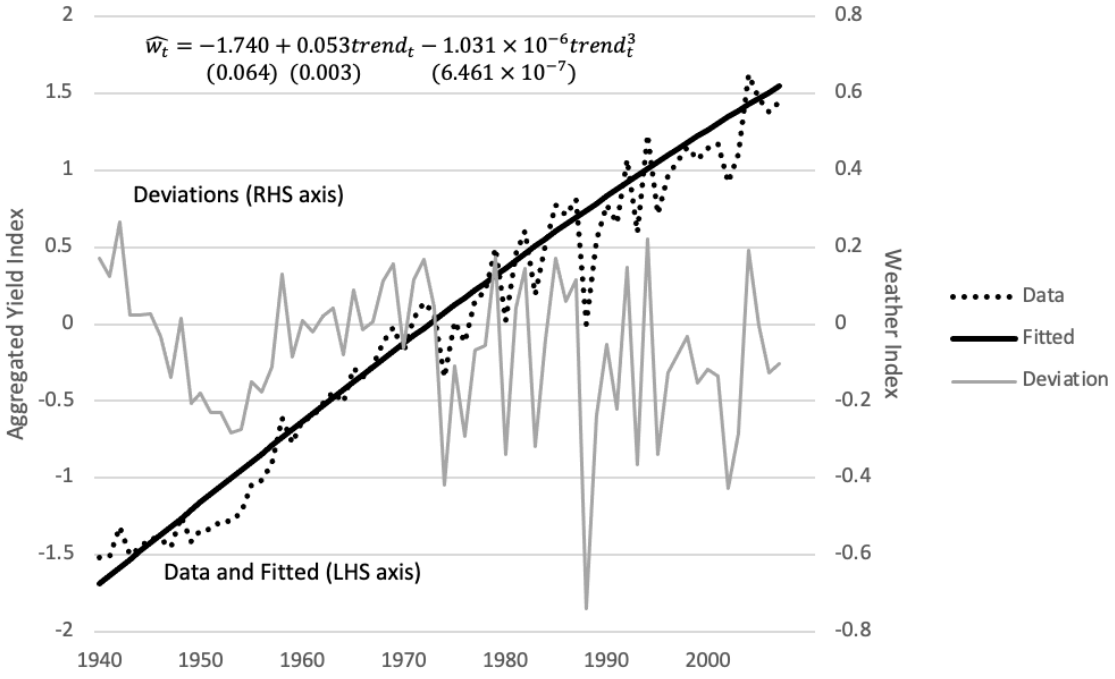
Figure 3.1: Inputs, Outputs and Multifactor Productivity, Logarithms, 1940–2007



Sources: University of Minnesota, InSTePP Center unpublished data. The SAES R&D series (excluding forestry) prior to 1980 is from USDA sources cited in Alston et al. (2010, appendix III) and for more recent years are compiled from unpublished USDA, CRIS data files. The USDA intramural series for years prior to 2001 are also from the USDA sources cited in Alston et al. (2010, appendix III) and NSF (various years) thereafter.

Notes: Public agricultural R&D includes SAES and USDA intramural spending, excluding forestry research. The series were deflated using an agricultural R&D deflator from InSTePP.

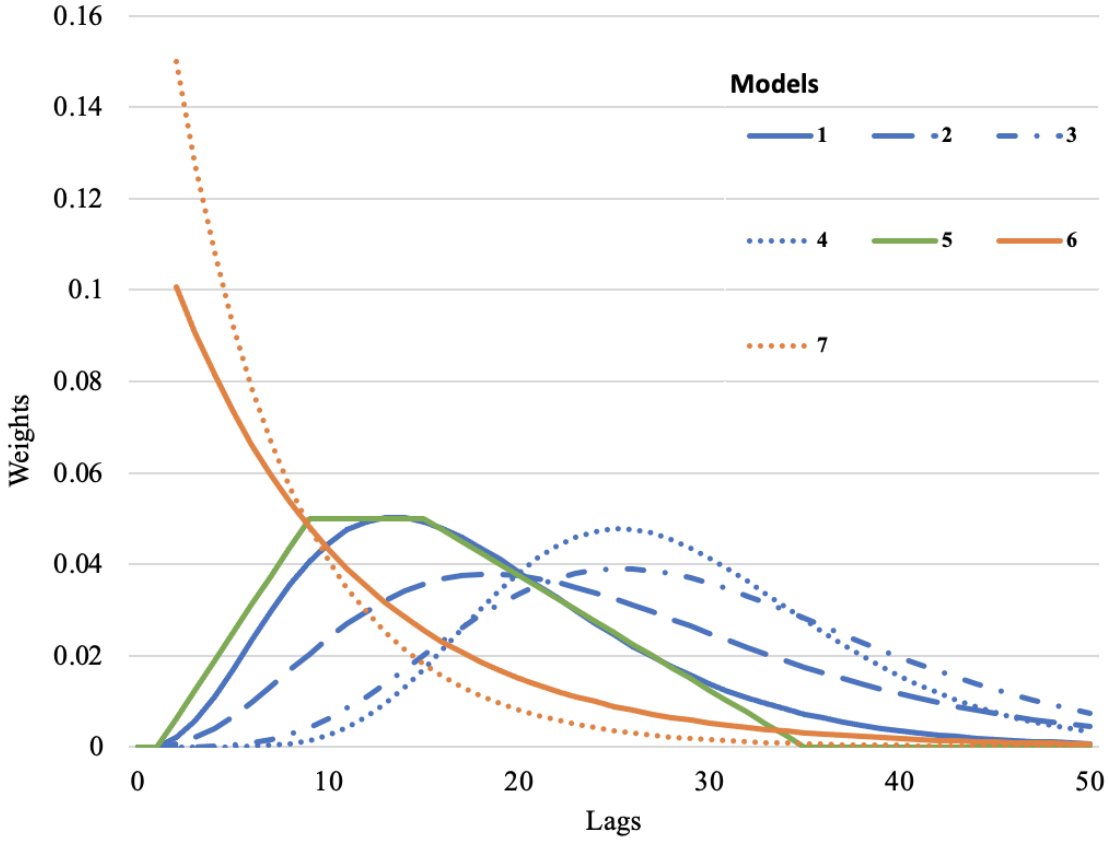
Figure 3.2: U.S Public Agricultural R&D, USDA Intramural and SAESs, 1890–2007



Sources: Developed by the authors.

Notes: The observed annual aggregated yields in year *t*, yield_t, were constructed as weighted averages of standardized annual yields of the top 10 field crops for the years 1940–2007. Each crop's annual share of the total value of production was used as its weight. The equation presents the fitted (linear and cubic) time-trend regression with standard errors in parentheses under each of the point estimates. T_t is the time trend created by calendar year minus 1939. The agricultural weather index in year *t* is given by yield deviations from the fitted value: $yield_t - \widehat{yield_t}$.

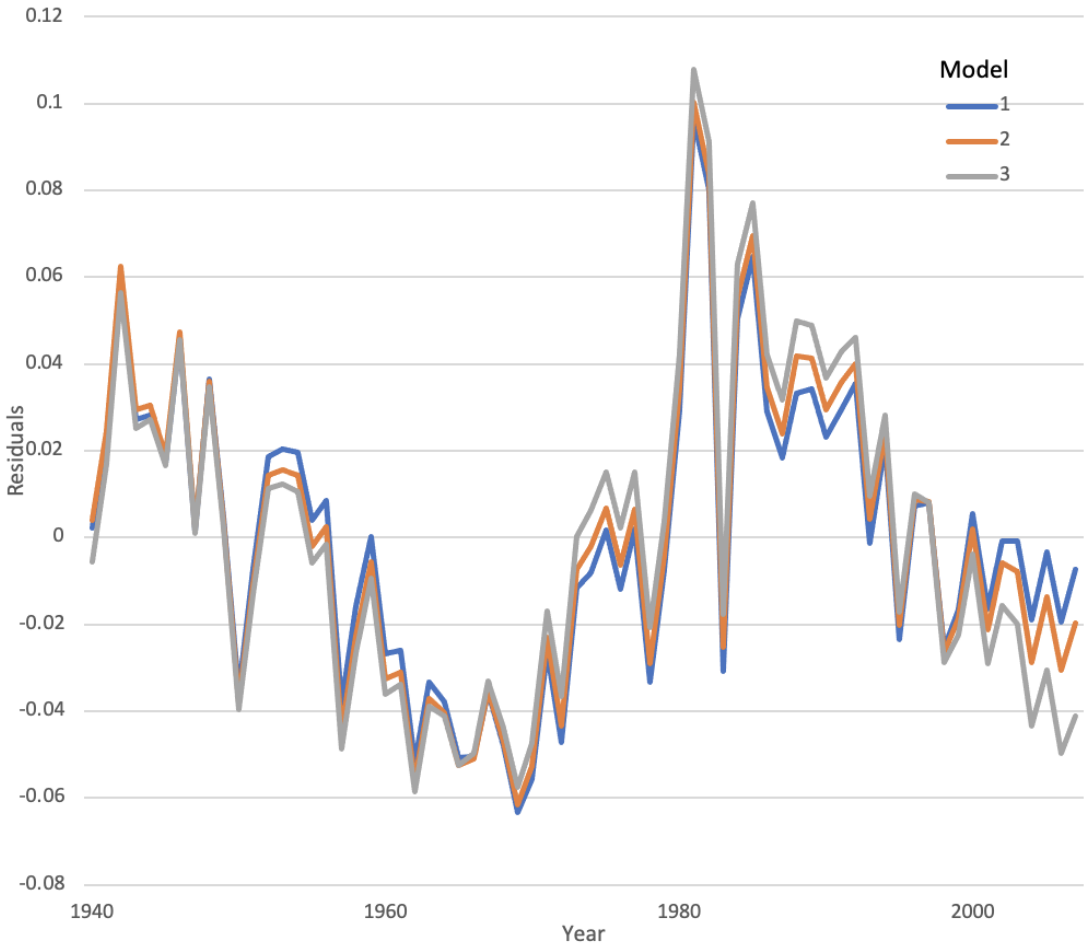
Figure 3.3: Fitted and Observed Composite Crop Yield Index, 1940–2007



Sources: Developed by the authors.

Notes: Gamma lag distribution models (Models 1–4) are shown in blue; the trapezoidal lag distribution (Model 5) is shown in green; the geometric lag distribution models (Models 6 and 7) are shown in orange; the Romer-Bloom model (Model 8) is not depicted here. Table 4 includes a summary of the parametrizations of these lag distribution models.

Figure 3.4: Lag Distribution Shapes for Models 1–7



Sources: Developed by the authors.

Notes: Derived from the fitted models estimated using OLS. Model numbers correspond to the first three gamma lag models in Table 3. See Appendix Table D for detailed regression results.

Figure 3.5: Residuals from the Models that Passed the Time-Series Tests (Models 1–3)

3.9 Tables

Table 3.1: Parameterization of Knowledge Stocks for the Alternative Models

Distributions	Parameters	Weights
Gamma	$\gamma, \lambda \in$ $\{x \times 0.05 + 0.6 \mid x \in Z, 0 \leq x \leq 7\}$, 64 combinations of γ and λ.	$b_k = \frac{(k-g+1)^{\gamma/\{(1-\gamma)\}}\lambda^{(k-g)}}{\sum_{k=g+1}^{50}[(k-g+1)^{\gamma/(1-\gamma)}\lambda^{(k-g)}]}$ for $g < k \leq 50$; otherwise $b_k = 0$
Trapezoidal	a = 1, b = 9, c = 15, d = 35, two years of gestation lag, then weights increase linearly for seven years, then stay constant for six years, and finally decline linearly for 20 years.	$b'_k = \begin{cases} 0, k < a \text{ or } k > d \\ \frac{2}{(d+c-a-b)}\frac{k-a}{b-a}, a \leq k < b \\ \frac{2}{(d+c-a-b)}, b \leq k < c \\ \frac{2}{(d+c-a-b)}\frac{d-k}{d-c}, c \leq k \leq d \end{cases}$ $b_k = \frac{b'_k}{\sum_{k=0}^{50} b'_k}$
Geometric	$\delta = 0.10$ or 0.15	$b'_k = (1-\delta)^k$ for $g < k \leq 50$; otherwise $b_k{}' = 0$ $b_k = \frac{b'_k}{\sum_{k=0}^{50} b'_k}$

Sources: Developed by the authors.

Notes: A two-year gestation period is equivalent to $g = 1$.

Table 3.2: Tests for Nonstationary Time Series

Variable	*Optimal lag (years)*	*Estimated Tau statistic*	*Critical values of Tau*		
			1%	*5%*	*10%.*
$\ln(MFP_t)$	1	-2.43	-3.70	-3.13	-2.83
$\Delta \ln(MFP_t)$	1	-9.14	-3.71	-3.14	-2.84
W_t	1	-4.51	-3.70	-3.13	-2.83

Sources: Developed by the authors.

Notes: Results obtained using STATA 17 to conduct the GLS-ADF test. The optimal lag was determined using the minimum Schwarz (1978) information criterion (SIC).

Table 3.3: Cointegration Tests with Alternative Lag Distribution Models

Model (1)	*Model specification*	(2)	*SSE* (3)	*GLS-ADF* $\ln(K_t)$ (4)	*GLS-ADF* $\Delta \ln(K_t)$ (5)	*Phillips-Perron test* (6)	*Johansen test* (7)
1	Gamma	$\gamma = 0.75$, $\lambda = 0.80$	0.074	Fail	10%	1%	Pass
2		$\gamma = 0.75$, $\lambda = 0.85$	0.083	Fail	1%	1%	Pass
3		$\gamma = 0.85$, $\lambda = 0.80$	0.096	Fail	10%	1%	Pass
4		$\gamma = 0.90$, $\lambda = 0.70$	0.098	Fail	Fail	1%	Pass
5	Trapezoidal		0.073	10%	Fail	1%	Fail
6	Geometric	$\delta = 0.10$	0.077	Fail	10%	Fail	Pass
7		$\delta = 0.15$	0.078	Fail	5%	Fail	Pass
8	Romer-Bloom	K_t	0.100	1%	Fail	Fail	Pass
9		$\ln(K_t)$	0.086	Fail	10%	1%	Pass

Sources: Developed by the authors.

Notes: SSE (sum of squared errors) is calculated from estimating equation (8) for models 1–7 and 9, and equation (8') for Model 8. In the case of the Romer-Bloom model (Model 8), the entries in columns (4) and (5) refer to K_t and ΔK_t rather than $\ln(K_t)$ and $\Delta \ln(K_t)$. The numbers in columns (4) through (7) indicate we reject the null hypothesis at the specific percentage significance levels shown (i.e., 1%, 5%, or 10%).

Table 3.4: Peak Lag Year and Mean Lag for Models 1–7

Model	*Model specification*		*Lag (years)*		
			Peak	*Mean*	*Maximum*
1	Gamma	$\gamma = 0.75, \lambda = 0.80$	13	17.8	50
2		$\gamma = 0.75, \lambda = 0.85$	18	23.3	50
3		$\gamma = 0.85, \lambda = 0.80$	25	28.3	50
4		$\gamma = 0.90, \lambda = 0.70$	25	27.6	50
5	Trapezoidal		9	15.7	35
6	Geometric	$\delta = 0.10$	2	10.7	∞
7		$\delta = 0.15$	2	7.7	∞

Sources: Developed by the authors.

Note: Derived from fitted models reported in Table 3.6

Table 3.5: Tests for Properties of Residuals from OLS Estimates of Model 1

Issue	*Test*	*Test statistic*	*P value*
Heteroskedasticity			
	White	$\chi^2_{(2)} = 6.22$	0.045
	Breusch-Pagan	$\chi^2_{(1)} = 0.05$	0.824
Autocorrelation			
First-order	Durbin-Watson	DW = 0.77	0.000
	Breusch-Godfrey	$\chi^2_{(1)} = 25.86$	0.000
Second-order	Breusch-Godfrey	$\chi^2_{(2)} = 26.41$	0.000
Third-order	Breusch-Godfrey	$\chi^2_{(3)} = 29.88$	0.000

Sources: Developed by the authors.

Notes: P value for the Durbin-Watson test is calculated using a normal approximation with mean and variance of the Durbin-Watson test statistic. Details can be found in 'lmtest' package in R.

Table 3.6: Dynamic OLS Regressions of MFP against Alternative Knowledge Stocks

Model	*Lag model (parameters)*	*Regressors*								
		Constant	$\ln(K_t)$	W_t	T_t	$\Delta\ln(K_t)$	$\Delta\ln(K_{t-1})$	$\Delta\ln(K_{t+1})$	$\Delta\ln(K_{t+2})$	$\Delta\ln(K_{t+3})$
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	Gamma (0.75, 0.80)	2.914*** (0.639)	0.277** (0.106)	0.020*** (0.006)	0.010** (0.004)	-8.200 (6.608)	9.156 (6.256)			
2	Gamma (0.75, 0.85)	2.874*** (0.654)	0.260** (0.109)	0.020*** (0.006)	0.012*** (0.004)	-12.334 (12.319)	17.354 (11.676)			
3	Gamma (0.85, 0.80)	1.922*** (0.558)	0.386*** (0.084)	0.025*** (0.005)	0.009*** (0.003)	2.224 (9.682)	7.225 (9.969)			
4	Gamma (0.90, 0.70)	2.455*** (0.371)	0.267*** (0.058)	0.026*** (0.003)	0.015*** (0.002)	88.593*** (22.111)		-72.038 (45.065)	-102.719** (45.755)	99.492*** (22.739)
5	Trapezoidal	2.843*** (0.658)	0.292*** (0.109)	0.020*** (0.005)	0.009** (0.004)	9.489* (5.604)		-9.561* (5.663)		
6	Geometric ($\delta = 0.10$)	3.235*** (0.352)	0.232*** (0.053)	0.019*** (0.005)	0.011*** (0.002)	-0.206 (0.801)		-0.846 (0.812)		
7	Geometric ($\delta = 0.15$)	3.414*** (0.309)	0.201*** (0.045)	0.018*** (0.005)	0.012*** (0.001)	-0.256 (0.575)		-0.546 (0.581)		
8	Romer–Bloom (K_t)	4.703*** (0.018)	7.29E-07* (4.08E-07) [0.082]	0.018*** (0.006)	0.007** (0.003)	6.28E-05 (3.96E-05)	3.45E-05 (3.85E-05)			
9	Romer–Bloom [$\ln(K_t)$]	1.415 (1.039)	0.351*** (0.115)	0.018*** (0.005)	0.004 (0.005)	-0.645 (3.693)		-4.168 (4.032)		

Sources: Developed by the authors.

Notes: Results using OLS and the Cochrane-Orcutt method are reported in the appendix. The gamma model parameters in parentheses are (γ, λ). Newey-West heteroskedasticity and autocorrelation consistent standard errors in parentheses in columns (1) through (5). ***, **, and * denote significance levels of 0.01, 0.05 and 0.10, respectively. Coefficients in column (3) are elasticities of MFP with respect to the knowledge stock. In the Romer Bloom model (Model 8) the elasticity is shown in square brackets in column (4), calculated at the median of constructed Romer-Bloom knowledge stock across the period 1940–2007. For Model 8, regressors in column (2) and (6) to (10) should be K_t and $\Delta\ln(Kt)$, as formulated in equation (8')

Table 3.7: Estimated Elasticities from Alternative Models and Estimators

Model	*Lag Model (parameters)*	*Elasticity of MFP with respect to K*			
		OLS	*Cochrane-Orcutt* GLS	*Paris-Winsten GLS*	*Dynamic OLS*
		(1)	(2)	(3)	(4)
1	Gamma (0.75, 0.80)	0.290*** (0.084)	0.306*** (0.109)	0.307*** (0.084)	0.277** (0.106)
2	Gamma (0.75, 0.85)	0.271** (0.130)	0.317** (0.143)	0.304*** (0.092)	0.260** (0.109)
3	Gamma (0.85, 0.80)	0.184 (0.374)	0.248 (0.185)	0.247** (0.094)	0.386*** (0.084)
4	Gamma (0.90, 0.70)	0.167 (0.466)	0.221 (0.192)	0.235** (0.098)	0.267*** (0.058)
5	Trapezoidal	0.282*** (0.078)	0.292*** (0.105)	0.299*** (0.084)	0.292*** (0.109)
6	Geometric ($\delta = 0.10$)	0.203*** (0.073)	0.212** (0.101)	0.227** (0.087)	0.232*** (0.053)
7	Geometric ($\delta = 0.15$)	0.183** (0.070)	0.190* (0.096)	0.205** (0.084)	0.201*** (0.045)
8	Romer–Bloom (K_t)	-0.077 (0.266)	-0.083 (0.093)	-0.115* (0.059)	0.082* (0.046)
9	Romer–Bloom [$\ln(K_t)$]	0.247* (0.144)	0.266* (0.150)	0.290** (0.116)	0.351*** (0.115)

Sources: Developed by the authors.

Notes: The gamma model parameters in parentheses are (γ, λ). Eicker-Huber-White (for Cochrane-Orcutt and Paris-Winsten GLS estimators) and Newey-West heteroskedasticity and autocorrelation consistent (for OLS and dynamic OLS) standard errors in parentheses in columns (1) through (3). ***, **, and * denote significance levels of 0.01, 0.05 and 0.10, respectively. Elasticities for models 1 to 7 and 9 are just point estimates of β_1 in equation (8). In the Romer-Bloom level model the elasticity is calculated as $\beta_1 K_t$ using the point estimate of β_1 in equation (8') with K_t as the median of the constructed Romer-Bloom knowledge stock across the period 1940–2007.

Table 3.8: Benefit-Cost Ratios from Various Models

Model	*Lag Model (Parameters)*	*Mean Lag (years)*	*OLS* (1)	*Cochrane-Orcutt* (2)	*Prais-Winsten* (3)	*Dynamic OLS* (4)
1	Gamma (0.75, 0.80)	17.8	**24.5** [10.3, 38.8]	**25.9** [7.4, 44.3]	**26.0** [11.8, 40.1]	**23.4** [5.5, 41.3]
2	Gamma (0.75, 0.85)	23.3	**21.4** [0.9, 42.0]	**25.0** [2.5, 47.6]	**24.0** [9.5, 38.5]	**20.5** [3.3, 37.8]
3	Gamma (0.85, 0.80)	28.3	13.0 [-39.8, 65.8]	17.5 [-8.7, 43.8]	**17.5** [4.2, 30.7]	**27.3** [15.4, 39.2]
4	Gamma (0.90, 0.70)	27.6	11.8 [-54.1, 77.8]	15.6 [-11.5, 42.8]	**16.6** [2.7, 30.5]	**18.9** [10.7, 27.0]
5	Trapezoidal	15.7	**24.4** [11.0, 37.7]	**25.2** [7.1, 43.3]	**25.8** [11.3, 40.4]	**25.2** [6.5, 44.0]
6	Geometric (0.10)	10.7	**18.4** [5.2, 31.6]	**19.2** [0.89, 37.5]	**20.5** [4.8, 36.3]	**21.0** [11.5, 30.5]
7	Geometric (0.15)	7.7	**16.9** [4.1, 29.6]	17.5 [-0.15, 35.1]	**18.9** [3.4, 34.3]	**18.5** [10.2, 26.8]
8	Romer-Bloom (K_t)		-5.7 [-45.1, 33.6]	-6.1 [-19.9, 7.7]	-8.5 [-17.3, 0.3]	6.1 [-0.7, 12.9]
9	Romer-Bloom [$\ln(K_t)$]		18.5 [-3.1, 40.0]	20.0 [-2.6, 42.5]	**21.7** [4.4, 39.1]	**26.3** [9.1, 43.6]

Sources: Developed by the authors.

Notes: The gamma model parameters in parentheses are (γ, λ). Entries in the table are the marginal benefit-cost ratios (BCRs) for an incremental investment in 1957 calculated using equation 12 and elasticities from Table 7, and a real discount rate of 3 percent per year. Numbers in square brackets are the upper and lower bounds of the 95 percent confidence interval for the BCR, and BCRs in bold are statistically significantly different from zero since their respective confidence intervals do not include zero.

Table 3.9: Effects of the Discount Rate on the Benefit-Cost Ratios from the Different Lag Models

Model	*Lag Model (Parameters)*	*Mean Lag (years)*	*Real Discount Rate*			
			$r = 0.001$	$r = 0.03$	$r = 0.05$	$r = 0.10$
			(1)	(2)	(3)	(4)
1	Gamma (0.75, 0.80)	17.8	**36.3** [8.6, 64.1]	**23.4** [5.5, 41.3]	**17.8** [4.2, 31.5]	**9.8** [2.3, 17.3]
2	Gamma (0.75, 0.85)	23.3	**35.5** [5.7, 65.3]	**20.5** [3.3, 37.8]	**14.7** [2.4, 27.0]	**7.1** [1.1, 13.1]
3	Gamma (0.85, 0.80)	28.3	**54.1** [30.6, 77.7]	**27.3** [15.4, 39.2]	**17.8** [10.1, 25.6]	**6.9** [3.9, 9.9]
4	Gamma (0.90, 0.70)	27.6	**37.6** [21.4, 53.9]	**18.9** [10.7, 27.0]	**12.1** [6.9, 17.4]	**4.5** [2.5, 6.4]
5	Trapezoidal	15.7	**37.6** [9.7, 65.6]	**25.2** [6.5, 44.0]	**19.6** [5.0, 34.2]	**11.4** [2.9, 19.8]
6	Geometric (0.10)	10.7	**26.9** [14.7, 39.1]	**21.0** [11.5, 30.5]	**18.1** [9.9, 26.4]	**13.3** [7.3, 19.3]
7	Geometric (0.15)	7.7	**22.4** [12.3, 32.4]	**18.5** [10.2, 26.8]	**16.5** [9.1, 23.8]	**12.8** [7.0, 18.5]
8	Romer-Bloom (K_t)		11.7 [-1.4, 24.8]	6.1 [-0.7, 12.9]	4.3 [-0.5, 9.1]	2.3 [-0.3, 4.9]
9	Romer-Bloom [$\ln(K_t)$]		**42.1** [14.5, 69.8]	**26.3** [9.1, 43.6]	**20.6** [7.1, 34.0]	**13.2** [4.5, 21.8]

Sources: Developed by the authors.

Notes: The gamma model parameters in parentheses are (γ, λ). Entries in the table are the marginal benefit-cost ratios (BCRs) for an incremental investment in 1957 calculated using equation 12 and elasticities from Table 6 (i.e., using the dynamic OLS estimators). Numbers in square brackets are the upper and lower bounds of the 95 percent confidence interval for the BCR, and BCRs in bold are statistically significantly different from zero since their respective confidence intervals do not include zero.

3.A Additional Tables

Table 3.A.1: Regression Results for Alternative Time Trend Models

	Models			
	Linear	*Quadratic*	*Cubic*	*Cubic*
	(1)	(2)	(3)	(4)
T_t	0.049***	0.056***	0.053***	0.045***
	(0.001)	(0.005)	(0.003)	(0.012)
T_t^2		-9.88E-05		2.96E-04
		(6.80E-05)		(4.15E-04)
T_t^3			-1.03E-06	-3.81E-06
			(6.46E-07)	(3.95E-06)
Constant	-1.672***	-1.752***	-1.740***	-1.686***
	(0.048)	(0.072)	(0.064)	(0.099)
Observations	68	68	68	68
R^2	0.961	0.962	0.963	0.963
Adjusted R^2	0.961	0.961	0.962	0.961
AIC	-25.510	-25.681	-26.124	-24.662

Sources: Developed by the authors.

Notes: The dependent variable is the annual weighted average yield of 10 field crops described in the data section of the main text. T_t is the time trend created by subtracting calendar year by 1939. ***, **, and * denote significance levels of 0.01, 0.05 and 0.10, respectively.

Table 3.A.2: Stationarity Tests for Knowledge Stocks from Alternative Models (Dickey-Fuller GLS Test)

Models	*Lag Model (Parameters)*	*DF-GLS on levels*					*DF-GLS on first differences*				
		lags	statistics	1%	5%	10%	lags	statistics	1%	5%	10%
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	Gamma (0.75, 0.80)	8	-2.500	-3.702	-2.839	-2.560	5	-2.766	-3.705	-2.982	-2.696
2	Gamma (0.75, 0.85)	5	-2.079	-3.702	-2.981	-2.695	1	-3.753	-3.705	-3.136	-2.837
3	Gamma (0.85, 0.80)	3	0.554	-3.702	-3.064	-2.772	2	-2.902	-3.705	-3.104	-2.809
4	Gamma (0.90, 0.70)	5	-0.640	-3.702	-2.981	-2.695	4	-1.683	-3.705	-3.026	-2.737
5	Trapezoidal	5	-2.822	-3.702	-2.981	-2.695	1	-2.654	-3.705	-3.136	-2.837
6	Geometric (0.10)	3	-0.762	-3.702	-3.064	-2.772	2	-3.015	-3.705	-3.104	-2.809
7	Geometric (0.15)	3	-0.629	-3.702	-3.064	-2.772	2	-3.301	-3.705	-3.104	-2.809
8	Romer-Bloom (K_t)	1	-4.613	-3.702	-3.131	-2.833	2	-1.377	-3.705	-3.104	-2.809
9	Romer-Bloom [$\ln(K_t)$]	1	-1.948	-3.702	-3.131	-2.833	2	-2.883	-3.705	-3.104	-2.809

Sources: Developed by the authors.

Notes: The Dicky-Fuller GLS is used to test the order of integration of knowledge stock $\ln(K_t)$ (or K_t for the Romer-Bloom level model) and its first differences $\Delta\ln(K_t)$ (or ΔK_t). The null hypothesis is that there exists a unit root in the tested time series, which implies the time series is nonstationary. The optimal lag length was chosen using the STATA 17 default minimum Schwarz (1978) information criterion (SIC). Critical values for significance levels (1%, 5%, and 10%) are listed following the test statistics.

Table 3.A.3: Cointegration Tests for Knowledge Stocks and MFP (Phillips-Perron Test)

Models	*Lag Model (Parameters)*	*Phillips-Perron Tests*				
		lags	statistics	1%	5%	10%
		(1)	(2)	(3)	(4)	(5)
1	Gamma (0.75, 0.80)	3	-4.873	-4.113	-3.483	-3.170
2	Gamma (0.75, 0.85)	3	-4.812	-4.113	-3.483	-3.170
3	Gamma (0.85, 0.80)	3	-4.276	-4.113	-3.483	-3.170
4	Gamma (0.90, 0.70)	3	-4.151	-4.113	-3.483	-3.170
5	Trapezoidal	3	-4.715	-4.113	-3.483	-3.170
6	Geometric (0.10)	3	-3.116	-4.113	-3.483	-3.170
7	Geometric (0.15)	3	-2.556	-4.113	-3.483	-3.170
8	Romer-Bloom (K_t)	3	-1.427	-4.113	-3.483	-3.170
9	Romer-Bloom [$\ln(K_t)$]	3	-4.866	-4.113	-3.483	-3.170

Sources: Developed by the authors.

Notes: The Phillips-Perron test is used to examine the cointegration relationship between $\ln(MFP_t)$ and $\ln(K_t)$ (or K_t for the Romer-Bloom level model). The null hypothesis is that the residual of regressing $\ln(MFP_t)$ on $\ln(K_t)$ (or K_t) contains a unit root (i.e., non-stationary). We imposed minimal restrictions by allowing the residual has a random walk, with or without drift, under the null hypothesis. Newey-West lags (lags=3) are used in the tests. Critical values for significance levels (1%, 5%, and 10%) are listed following the test statistics.

Table 3.A.4: Cointegration Tests for Knowledge Stocks and MFP (Johansen test)

Models	*Lag Model (Parameters)*	*AIC*			*HQIC*			*SBIC*			*5% critical value*	
		lags	Trace statistics		lags	Trace statistics		lags	Trace statistics			
		(1)	Maximum rank 0 (2)	Maximum rank 1 (4)	(6)	Maximum rank 0 (7)	Maximum rank 1 (9)	(6)	Maximum rank 0 (7)	Maximum rank 1 (9)	Maximum rank 0 (10)	Maximum rank 1 (11)
1	Gamma (0.75, 0.80)	9	32.154	12.045	8	39.907	9.574	4	42.647	7.572	25.320	12.250
2	Gamma (0.75, 0.85)	6	32.693	14.144	6	32.693	14.144	3	32.797	10.646	25.320	12.250
3	Gamma (0.85, 0.80)	5	34.381	7.495	4	37.081	8.239	4	37.081	8.239	25.320	12.250
4	Gamma (0.90, 0.70)	6	25.400	8.633	4	63.194	9.381	3	58.206	12.093	25.320	12.250
5	Trapezoidal	3	19.204	7.542	3	19.204	7.542	3	19.204	7.542	25.320	12.250
6	Geometric (0.10)	5	27.346	5.947	4	20.444	6.309	2	22.864	6.187	25.320	12.250
7	Geometric (0.15)	5	26.142	5.901	5	26.142	5.901	2	21.040	4.831	25.320	12.250
8	Romer-Bloom (K_t)	5	40.901	8.916	4	32.712	6.839	2	38.815	13.526	25.320	12.250
9	Romer-Bloom [$\ln(K_t)$]	6	26.380	5.420	2	35.520	9.663	2	35.520	9.663	25.320	12.250

Sources: Developed by the authors.

Notes: The Johansen cointegration test is used to examine the cointegration relationship between $\ln(MFP_t)$ and $\ln(K_t)$ (or K_t for the Romer-Bloom level model). Maximum rank 0 (or 1) represents the case where the maximum rank of the cointegration matrix is 0 (or 1). The null hypothesis is that there exists up to r cointegration relations, where r starts from 0, then 1, and so on. In our specification, since there are two time-series, we perform Johannsen tests from $r = 0$ to $r = 1$. If according to the trace statistics, we reject the null that $r = 0$ but fail to reject $r = 1$, this means there only exists one cointegration relationship between $\ln(MFP_t)$ and $\ln(K_t)$ (or K_t). The 5% critical values for rank 0 and rank 1 are presented in columns (10) and (11). The optimal lags are selected by the Akaike Information Criterion (AIC), the Hannan-Quinn Information Criterion (HQIC), and the Schwarz-Bayesian Information Criterion (SBIC).

Table 3.A.5: Heteroskedasticity Test

Models	*Lag Model*	*Breusch-Pagan*		*White*	
	(Parameters)	$\chi^2_{(1)}$	P value	$\chi^2_{(2)}$	P value
1	Gamma (0.75, 0.80)	0.050	0.824	6.219	0.045
2	Gamma (0.75, 0.85)	0.003	0.959	4.691	0.096
3	Gamma (0.85, 0.80)	1.212	0.271	3.939	0.140
4	Gamma (0.90, 0.70)	1.622	0.203	3.914	0.141
5	Trapezoidal	0.082	0.775	6.512	0.039
6	Geometric (0.10)	0.009	0.924	6.993	0.030
7	Geometric (0.15)	0.003	0.957	6.742	0.034
8	Romer-Bloom (K_t)	2.877	0.090	4.897	0.086
9	Romer-Bloom [$\ln(K_t)$]	0.092	0.762	5.317	0.070

Sources: Developed by the authors.

Notes: The Breusch-Pagan and White tests are chosen to test the heteroskedasticity of error terms in equations (8) and (8'). These two tests examine linear and non-linear heteroskedasticity, respectively.

Table 3.A.6: Autocorrelation Test

Models	*Lag Model (Parameters)*	*Durbin-Watson*		*Breusch-Godfrey*					
		First order		*First order*		*Second order*		*Third order*	
		DW statistics	P value	$\chi^2_{(1)}$	P value	$\chi^2_{(2)}$	P value	$\chi^2_{(3)}$	P value
1	Gamma (0.75, 0.80)	0.769	0.000	25.859	0.000	26.410	0.000	29.876	0.000
2	Gamma (0.75, 0.85)	0.677	0.000	29.671	0.000	30.233	0.000	33.732	0.000
3	Gamma (0.85, 0.80)	0.590	0.000	33.599	0.000	34.259	0.000	37.681	0.000
4	Gamma (0.90, 0.70)	0.580	0.000	34.085	0.000	34.737	0.000	38.092	0.000
5	Trapezoidal	0.790	0.000	25.066	0.000	25.705	0.000	29.288	0.000
6	Geometric (0.10)	0.797	0.000	25.321	0.000	26.577	0.000	30.652	0.000
7	Geometric (0.15)	0.799	0.000	25.337	0.000	26.704	0.000	30.860	0.000
8	Romer-Bloom (K_t)	0.591	0.000	33.767	0.000	34.720	0.000	38.381	0.000
9	Romer-Bloom $[\ln(K_t)]$	0.696	0.000	29.298	0.000	30.339	0.000	34.047	0.000

Sources: Developed by the authors.

Notes: The Durbin-Watson and Breusch-Godfrey tests are chosen to test the autocorrelation of error terms in equations (8) and (8'). Test statistics and p values are presented. The results strongly indicate that the error term is at least first-order autocorrelated.

Table 3.A.7: OLS Regressions of MFP against Knowledge Stocks with Alternative Lag Models

Model	*Lag Model, Parameters*	*Regressors*				
		Constant	$\ln(K_t)$	K_t	W_t	T_t
		(1)	(2)	(3)	(4)	(5)
1	Gamma (0.75, 0.80)	2.882*** (0.555)	0.290*** (0.084)		0.020*** (0.006)	0.009*** (0.003)
2	Gamma (0.75, 0.85)	3.078*** (0.821)	0.271** (0.130)		0.021*** (0.006)	0.009* (0.005)
3	Gamma (0.85, 0.80)	3.683 (2.232)	0.184 (0.374)		0.021*** (0.007)	0.012 (0.015)
4	Gamma (0.90, 0.70)	3.781 (2.787)	0.167 (0.466)		0.021*** (0.007)	0.012 (0.019)
5	Trapezoidal	2.911*** (0.517)	0.282*** (0.078)		0.020*** (0.006)	0.009*** (0.002)
6	Geometric ($\delta = 0.10$)	3.385*** (0.500)	0.203*** (0.073)		0.018*** (0.006)	0.013*** (0.002)
7	Geometric ($\delta = 0.15$)	3.502*** (0.483)	0.183** (0.070)		0.017*** (0.005)	0.013*** (0.002)
8	Romer-Bloom (K_t)	4.784*** (0.024)		-6.84E-07 (2.35E-06) [-0.077]	0.019*** (0.006)	0.021** (0.008)
9	Romer-Bloom [$\ln(K_t)$]	2.253 (1.474)	0.247* (0.144)		0.019*** (0.006)	0.009* (0.005)

Sources: Developed by the authors.

Notes: Estimates obtained using OLS. The gamma model parameters in parentheses are (γ, λ). Newey-West heteroskedasticity and autocorrelation consistent errors (with pre-whitening and adjust for small sample) in parentheses in columns (1) through (5). ***, **, and * denote significance levels of 0.01, 0.05 and 0.10, respectively. Coefficients in column (3) are elasticities of MFP with respect to the knowledge stock. For the Romer-Bloom model (Model 8) the elasticity is shown in square brackets in column (4), calculated at the median of constructed Romer-Bloom knowledge stock across the period 1940–2007.

Table 3.A.8: Cochrane-Orcutt Regressions of MFP against Knowledge Stocks with Alternative Lag Models

Model	*Lag Model, Parameters*	*Regressors*				
		Constant	$\ln(K_t)$	K_t	W_t	T_t
		(1)	(2)	(3)	(4)	(5)
1	Gamma (0.75, 0.80)	2.781*** (0.725)	0.306*** (0.109)		0.026*** (0.004)	0.008** (0.003)
2	Gamma (0.75, 0.85)	2.788*** (0.914)	0.317** (0.143)		0.026*** (0.004)	0.007 (0.005)
3	Gamma (0.85, 0.80)	3.297*** (1.135)	0.248 (0.185)		0.026*** (0.004)	0.009 (0.007)
4	Gamma (0.90, 0.70)	3.460*** (1.178)	0.221 (0.192)		0.026*** (0.004)	0.010 (0.007)
5	Trapezoidal	2.845*** (0.705)	0.292*** (0.105)		0.026*** (0.004)	0.009*** (0.003)
6	Geometric ($\delta = 0.10$)	3.331*** (0.707)	0.212** (0.101)		0.026*** (0.004)	0.012*** (0.003)
7	Geometric ($\delta = 0.15$)	3.461*** (0.682)	0.190* (0.096)		0.026*** (0.004)	0.013*** (0.002)
8	Romer–Bloom (K_t)	4.794*** (0.034)		-7.30E-07 (8.26E-07) [-0.083]	0.026*** (0.004)	0.021*** (0.004)
9	Romer-Bloom [$\ln(K_t)$]	2.055 (1.557)	0.266* (0.150)		0.026*** (0.004)	0.009 (0.005)

Sources: Developed by the authors.

Notes: Estimates obtained using the Cochrane-Orcutt method. The gamma model parameters in parentheses are (γ, λ). Eicker-Huber-White standard errors in parentheses in columns (1) through (5). ***, **, and * denote significance levels of 0.01, 0.05 and 0.10, respectively. Coefficients in column (3) are elasticities of MFP with respect to the knowledge stock. For the Romer-Bloom model (Model 8) the elasticity is shown in square brackets in column (4), calculated at the median of constructed Romer-Bloom knowledge stock across the period 1940–2007.

Table 3.A.9: Prais-Winsten Regressions of MFP against Knowledge Stocks with Alternative Lag Models

Model	*Lag model (parameters)*	*Regressors*				
		Constant	$\ln(K_t)$	K_t	W_t	T_t
		(1)	(2)	(3)	(4)	(5)
1	Gamma (0.75, 0.80)	2.772*** (0.550)	0.307*** (0.084)		0.026*** (0.004)	0.008*** (0.003)
2	Gamma (0.75, 0.85)	2.873*** (0.577)	0.304*** (0.092)		0.026*** (0.004)	0.008** (0.003)
3	Gamma (0.85, 0.80)	3.304*** (0.559)	0.247** (0.094)		0.026*** (0.004)	0.009** (0.004)
4	Gamma (0.90, 0.70)	3.373*** (0.586)	0.235** (0.098)		0.026*** (0.004)	0.009** (0.004)
5	Trapezoidal	2.798*** (0.561)	0.299*** (0.084)		0.026*** (0.004)	0.009*** (0.003)
6	Geometric ($\delta = 0.10$)	3.222*** (0.603)	0.227** (0.087)		0.026*** (0.004)	0.013*** (0.002)
7	Geometric ($\delta = 0.15$)	3.348*** (0.592)	0.205** (0.084)		0.026*** (0.004)	0.013*** (0.002)
8	Romer-Bloom (K_t)	4.781*** (0.018)		-1.02E-06 (5.26E-07) [-0.115]	0.026*** (0.004)	0.023*** (0.002)
9	Romer-Bloom [$\ln(K_t)$]	1.809 (1.191)	0.290** (0.116)		0.026*** (0.004)	0.008* (0.004)

Sources: Developed by the authors.

Notes: Results using OLS and the Cochrane-Orcutt method are reported in the appendix. The gamma model parameters in parentheses are (γ, λ). Eicker-Huber-White standard errors in parentheses in columns (1) through (5). ***, **, and * denote significance levels of 0.01, 0.05 and 0.10, respectively. Coefficients in column (3) are elasticities of MFP with respect to the knowledge stock. In the Romer-Bloom model (Model 8) the elasticity is shown in square brackets in column (4), calculated at the median of constructed Romer-Bloom knowledge stock across the period 1940–2007.