Estimation of the costs and benefits of climate change mitigation

by

Kendon Matthew Bell

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Maximilian Auffhammer, Chair
Professor David Zilberman
Associate Professor Solomon Hsiang
Professor Peter Berck

Spring 2017

Estimation of the costs and benefits of climate change mitigation

Abstract

Estimation of the costs and benefits of climate change mitigation

by

Kendon Matthew Bell

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Maximilian Auffhammer, Chair

The global climate is changing, and it is incumbent on researchers to determine both the costs and benefits of slowing this change. This dissertation contributes to the study of both of these goals. One chapter examines the effect that measurement error in control variables can have on the empirical estimation of the relationship between economic variables and temperature, using the case of maize in the United States of America. This chapter suggests that measurement error in precipitation could bias temperature coefficients away from zero, in this context. One chapter studies the effect of weather variables on the world's largest livestock industry, dairy, in the New Zealand context, and finds large and different effects of weather in summer and winter, and finds that bottom-line conclusions are sensitive to the assumption that the effects do not vary by season. Finally, with a co-author, one chapter studies the potential for a low carbon fuel source that uses a proven, mature, technology, sugarcane ethanol, and finds that a substantial proportion of carbon-intensive fuel could be displaced, under modest subsidies, suggesting that some fossil fuels can be replaced with low abatement costs.

Contents

List of Figures

List of Tables

Acknowledgments

I want to thank my advisor, Professor Maximilian Auffhammer, for his encouragement, critiques, and pragmatism. I would also like to extend special thanks to Professor David Zilberman for his encouragement, advice, and mentorship, throughout my five years at Berkeley. I also thank David Zilberman for inviting me to share our co-authored research in this dissertation. I thank Professor Solomon Hsiang for his advice, support, critiques, and most importantly, inspiration. Professors Peter Berck, Brian Wright, and Michael Anderson were superb members of my oral and/or dissertation committees and I thank them graciously. I have also benefited greatly from both conversations with and inspiration from Professors Michael Roberts and Wolfram Schlenker.

I am grateful to the friends I've made during my studies, who helped with both work and play: Fiona Burlig, Louis Preonas, Andrew Stevens, James Gillan, Rebecca Taylor, David McLaughlin, Zhimin Li, Joshua Blonz, and Bret Strogen.

Finally, thanks to my parents, Linda Kelly and Philip Bell, who provided love and encouragement, my siblings Courtney and Jeremy Bell, who gave me friendship, love, and much-needed time away from my studies each year. I thank my partner, Emma, for her support, encouragement, love, and, most deservedly, patience.

Chapter 1

Overview

This dissertation brings together research on three topics in environmental and resource economics. Specifically, each of the chapters contributes to either the study of the costs or of the benefits of slowing climate change, which is an issue which global policy makers continue to consider.

The first chapter investigates the potential effect of measurement error in control variables on the estimated responses of economic variables to temperature. I use the case of agriculture, in which an extensive literature of process crop models provides a set of output response functions that is uncontaminated by errors in measurement. I show that temperature responses in statistically emulated maize models are attenuated when including precipitation, while the response in the empirical counterpart is not. I also find that the precipitation response functions in the emulated maize models conform to reasonable prior expectations, while the response in the empirical counterpart does not. I also find that the residual correlation between extreme temperature and precipitation is large and negative. The results are consistent with two narratives: one where the true precipitation response function is very different from those in agronomic crop models and one where mis-measurement of precipitation biases the empirical extreme temperature response away from zero. I provide suggestive evidence that using long differences can partially solve the problem and I propose a solution that uses instruments.

The second chapter examines the relationship between weather and dairy production in New Zealand, which is a major pasture-based livestock producer. The vast majority of the land used in the global agricultural system supports livestock systems, with the vast majority of this land in pasture. Due to concern about the environmental effects of changing land uses, and expected future demand increases for animal products, it is crucial to understand how these systems will react to future climate change. Using data on the production and quality of milk produced in New Zealand, this chapter estimates the nonlinear relationship between weather and dairy production. I estimate models both restricting response functions to be the same throughout the dairy season and allowing for heterogeneity by time-of-year. I find large and negative impacts of moderate to high temperatures in summer months and

large and positive impacts of moderate temperatures during winter months. I give suggestive evidence that allowing for seasonality in responses results in less pessimistic projections of the response to future climate change in this context. I find statistically, but not economically, significant negative impacts of rising temperatures on milk quality.

In the third chapter, which is joint work with David Zilberman, we develop a supply model for ethanol production in Brazil with spatially disaggregated potential yield, freight costs, and pasture land available for conversion. We show that, under the assumptions of free capital markets, constant prices, and a modest increase over the current oil-equivalent price, a non-trivial amount of future global liquid fossil fuel can be profitably displaced by Brazilian ethanol production using existing pasture land. Along with policies to encourage the intensification of existing beef production, the dominant current land use, this new production can occur without the use of additional agricultural land, assuaging concerns about indirect land use change. At the current ethanol price, which includes the subsidizing effect of the mandate, the model predicts a substantial expansion of sugarcane ethanol, indicating that real-world considerations, such as capital controls and institutional, policy, and price uncertainty, are considerable barriers to investment in this context.

Chapter 2

Does the 'Iron Law' always hold? The impact of measurement error in climate econometrics

2.1 Introduction

Accurate measurement of the effect of climate on societies is crucial to understanding the costs and benefits of climate change, and identifying opportunities for reducing economic dependence on the climate. In climate econometrics, researchers often focus on the response of temperature, while controlling for precipitation (Schlenker and Roberts 2009; Burke, Miguel, et al. 2009; Deschênes and Greenstone 2011; Deryugina and Hsiang 2014; Burke, Hsiang, and Miguel 2015). Generally, studies that find large temperature effects also find relatively small effects of precipitation. Precipitation has been shown to be difficult to accurately interpolate across space, while temperature interpolations are quite accurate (Hijmans et al. 2005; Lobell and Asseng 2017).

It is well known that classical measurement error in a single regressor results in bias towards zero, or *attenuation bias*, in OLS regression, for the estimated coefficient on that regressor; Hausman (2001) refers to this bias as the "Iron Law of Econometrics". However, less attention has been paid to the potential bias in coefficients of interest that might result from including control variables that are mis-measured (Atkinson and Crocker 1992). In general, the coefficients on all variables will be biased in this situation; Garber and Klepper (1980) provide a formula that shows that, when a single control variable is mis-measured, this bias depends on the correlations between all variables, and is not necessarily attenuating, except for the coefficient on the mis-measured variable. For research on weather impacts, if we believe that error in the measurement of precipitation is large relative to that for temperature, there is cause for concern about bias away from zero in the temperature coefficients (Lobell 2013). In the extreme, if precipitation was measured so poorly that there was no signal at

all, and temperature was measured correctly, then all of the effect of precipitation would load on to the temperature effect, if these variables are correlated.

This chapter provides several pieces of evidence that suggest that, for the case of maize in the United States of America, measurement error in precipitation does indeed cause bias away from zero for the effects of extreme hot temperature. Firstly, I statistically emulate[1] nonlinear temperature response functions for each of seven process crop models, using the exact inputs that forced the models, thus providing responses that are uncontaminated by measurement error. I then compare the effect of introducing precipitation as a control variable in the emulated process models to that for the equivalent empirical model and show that the effects of extreme temperature attenuate substantially in the process models, while that in the empirical model does not.[2] Next, I show that the process models produce inverted-U-shaped precipitation response functions, while the empirical model does not, consistent with attenuation bias in the precipitation coefficients in the empirical model. Next, I show that the residual correlation between precipitation and extreme hot temperatures, after partialling out fixed effects and other temperature variables, is large and negative, consistent with that which would be required for bias away from zero in the extreme hot temperature coefficient.

The results suggest that future changes in precipitation will have a larger impact than empirical maize models suggest, and future changes in temperature will have a smaller impact.

A natural question to ask is does this bias matter for climate *change* impacts estimation? For many studies, the bottom-line quantity of interest is not necessarily the impact of contemporaneous temperature, but a projection or forecast of the response of an economic variable under future expected changes in both temperature and precipitation. I derive an expression that shows that, when future changes in temperature and precipitation retain their historical bivariate relationship, coefficients on correctly measured temperature variables can successfully recover an unbiased estimate of the future change in an economic variable, conditional, of course, on the historical relationship between weather and the economic variable also continuing to hold. In essence, the projected future temperature impact subsumes the future precipitation impact. However, climate models generally do not project that the future bivariate relationship between temperature and precipitation will be the same as today, with temperature increasing everywhere and precipitation changes varying greatly across models and space, so this result is not necessarily cause for optimism.

Using long differences models (Burke and Emerick 2016) over periods of temperature change could partially alleviate the problem of measurement error in precipitation for two reasons. Firstly, the long differences model uses fewer time fixed effects, reducing the amount of signal that is taken out of the data by the fixed effects estimator. Secondly, the longer run temperature change *may* break the problematic correlation between extreme temperature

[1] See in Blanc and Sultan (2015) and Schauberger et al. (2017) for examples of this approach.

[2] Asymptotically, omitting a control variable causes the same bias as including a highly mis-measured proxy for the same control variable.

and precipitation. I find that, indeed, using this longer run variation reduces the residual correlation between extreme temperature and precipitation; however, the correlation remains large. I also find that moderate temperatures in the long differences model have a significant negative residual correlation with precipitation, while they do not in the fixed effects model; this finding then causes concern about negative bias in the effect of moderate temperatures.

The most optimistic path forward, for contexts in which researchers expect precipitation impacts to be large, and precipitation monitors are dense, is computing two interpolations of each grid cell with different sets of monitors for each grid cell-day, then instrument one interpolation with the other. However, this relies on correct spatial aggregation within each unit and time period as incorrect aggregation leaves relevant variables in the error term, and the weather at the incorrectly omitted grid cell is likely to be more correlated with one of the sets of monitors more than the other. A future iteration of this project will implement this solution.

This chapter proceeds as follows: the following section briefly outlines the theory of the effect of measurement error in control variables in OLS regression. Next, I outline the data I use in my analyses; following that, I specify the regression methods I use and discuss the results. Finally, I conclude with some thoughts about the implications of this work.

2.2 Theory

This section outlines the econometric theory of the effect of classical measurement error in one variable on the coefficient on a second variable. Griliches (1957), Theil (1957), Blomqvist (1972), McCallum (1972), Wickens (1972), Levi (1973), Aigner (1974) all made relevant contributions that analyzed some aspect of this problem. The exposition here draws heavily on Garber and Klepper (1980) and I express all variables in demeaned form.

Suppose we have an economic variable, y, that depends on both temperature, T^*, and precipitation, P^*:

$$y = \beta_T^* T^* + \beta_P^* P^* + \varepsilon \tag{2.1}$$

where ε is a standard noise term. Suppose further that precipitation is measured with error such that observed precipitation is:

$$P = P^* + u \tag{2.2}$$

Modifying the results presented in Garber and Klepper (1980)[3] provides the expressions for the probability limits on each estimated coefficient in a regression of y on the true

[3]These authors cite Chow (1957), which is not readily accessible, when they obtain these results.

temperature, T^*, and the mis-measured precipitation, P:

$$\hat{\beta}_P = \left(\frac{\sigma_P^*}{\sigma_P^* + \sigma_u^2}\right)\beta_P^* \tag{2.3}$$

$$\hat{\beta}_T = \beta_T^* + \left(\frac{\sigma_u^2}{\sigma_P^* + \sigma_u^2}\right)\gamma_{PT}^*\beta_P^* \tag{2.4}$$

where $\hat{\beta}_P$ and $\hat{\beta}_T$ are the estimated coefficients in the regression of y on T^* and P, σ_P^* is the variance in the true precipitation variable, σ_u^2 is the variance of the measurement error term u, β_P^* and β_T^* are the true effects of precipitation and temperature, and γ_{PT}^* is the coefficient in a regression of the true precipitation, P^*, on the true temperature, T^*.

Equation (2.3) shows the familiar "Iron Law" result, where the precipitation coefficient is attenuated. However, Equation (2.4), shows that the bias in the temperature coefficient, $\hat{\beta}_T - \beta_T^*$, depends on the measurement error to total variance ratio in precipitation, the linear relationship between precipitation and temperature, and the size of the precipitation effect in the main regression. The measurement error to total variance ratio is in the interval $[0, 1)$, so the total bias is bounded between 0 and the product of the coefficient in a regression of P^* on T^* and the effect of precipitation in the main estimating equation. Importantly, this implies that only control variables that are both relevant in the main relationship of interest and correlated with the variable of interest can cause substantial bias, and that the bias increases with the degree of mis-measurement.

If temperature and precipitation are negatively correlated, and precipitation has a positive effect, then the bias in the temperature coefficient is negative. If the true effect of temperature is negative or zero, then this bias is away from zero.

Climate change impacts

If the researcher is not interested in β_T^* and β_P^*, but only in the total impact of changes in both of these variables, Δy, as in the case of projections of the future impact of climate change, then the bias I described in the previous subsection does not necessarily imply biased estimates of Δy. The formal expression for the bias in $\widehat{\Delta y}$, ignoring error terms, is:

$$\begin{aligned}
\widehat{\Delta y} - \Delta y &= \left(\hat{\beta}_T - \beta_T^*\right)\Delta T^* + \left(\hat{\beta}_P - \beta_P^*\right)\Delta P^* \\
&= \left(\frac{\sigma_u^2}{\sigma_P^* + \sigma_u^2}\right)\gamma_{PT}^*\beta_P^*\Delta T^* - \left(\frac{\sigma_u^2}{\sigma_P^* + \sigma_u^2}\right)\beta_P^*\Delta P^* \\
&= \left(\gamma_{PT}^*\Delta T^* - \Delta P^*\right)\left(\frac{\sigma_u^2}{\sigma_P^* + \sigma_u^2}\right)\beta_P^*
\end{aligned} \tag{2.5}$$

Equation (2.5) states that projected changes in y will be unbiased if the relationship between temperature and precipitation in *future changes* is the same as it is in levels (i.e. $\gamma_{PT}^*\Delta T^* =$

ΔP^*). Intuitively, if climate change preserves the historical relationship between temperature and precipitation, and current precipitation effects are subsumed in temperature effects, then projected effects of temperature changes will also subsume the future effects of precipitation changes.

2.3 Data

Yield data

I collect historical maize yield data from USDANASS for the empirical model. Following Schauberger et al. (2017), I subset these data to include counties where 90% or more of the maize land is rain-fed, as measured in the MIRCA2000 dataset, which measures rain-fed and irrigated land uses by crop for the year 2000. The empirical model's time period is defined by the availability of the weather dataset I use.

The "Intersectoral Impact Model Intercomparison Project" (ISIMIP) provides modeled yields from seven crop process models: EPIC, GEPIC, IMAGE, LPJ-GUESS, LPJmL, pDSSAT and PEGASUS. These data are gridded at the $0.5°\times0.5°$ level and, also following Schauberger et al. (2017), and motivated by a desire to use the same geographic scale for both the empirical and the process models, I spatially aggregate these data up to the county level using area-overlap weights. The historical ISIMIP yield data are available for various periods within 1971-2010; these time periods define the process models' time periods.

Weather data

I use a reanalysis weather dataset, AgMERRA, for the empirical data.[4] Precipitation, relative humidity, and wind speed are gridded at the $0.25°\times0.25°$ level; temperature is gridded at the $0.5°\times0.5°$ level, and solar radiation is gridded at the $1°\times1°$ level.[5] Following Schlenker and Roberts (2009), I construct linear spline transformations of these variables at the grid cell-day level, spatially aggregate the variables using area-overlap weights, and temporally aggregate these using the sum over days in the March-August growing season. These data are available for the period from 1980-2010, defining the time period for the empirical model.

[4]See Auffhammer, Hsiang, et al. (2013) for a discussion of the use of reanalysis data in economic applications.

[5]I have two motivations for using these data instead of a higher spatial detail dataset, such as PRISM. Firstly, the Agricultural Model Intercomparison Project's (AgMIP) subproject, the Global Gridded Crop Model Intercomparison (GGCMI) uses these data to force the process crop models being compared. When the crop model output from those data are available (I have been told they will be released "soon"), they can be readily included in this project. Secondly, the AgMERRA data include relative humidity, wind speed, and solar radiation, which are variables that are also considered important for crop growth, along with temperature and precipitation.

The ISIMIP data also includes the exact weather used to force the process models. I transform and aggregate these data in the same manner as the empirical data. The ISIMIP data contains forced crop models from five different climate models, and I pool all of these in all the process model regressions. However, to ensure the average climate is similar in the process and empirical models, I subset to only counties that appear in the empirical sample.

2.4 Methods

Comparing regressions across process and empirical models

In the first stage of this analysis, I will compare the process and empirical models across several specifications with increasingly detailed controls. I can represent all the specifications in this stage in the following form:

$$y_{it} = f(\boldsymbol{T}_{it}) + g(\boldsymbol{P}_{it}) + \boldsymbol{z}_{it}\boldsymbol{\alpha} + \gamma_i + \delta_{i1}t + \delta_{i2}t^2 + \varepsilon_{it} \tag{2.6}$$

where y_{it} is the maize yield in county i and year t, f is a function of the vector of daily temperature maxima and minima within a county-year, g is a function of the vector of daily precipitation within a county-year, $\boldsymbol{z}_{it}$ is a vector of other control variables, γ_i is a county fixed effect, and $\delta_{i1}t + \delta_{i2}t^2$ is a state-specific quadratic trend.

The base specification is similar to the linear spline specification in Schlenker and Roberts (2009) with no control variables:

$$f(\boldsymbol{T}_{it}) = S^1_{0,27}(\boldsymbol{T}_{it}) \tag{2.7}$$

$$g(\boldsymbol{P}_{it}) = 0 \tag{2.8}$$

$$\boldsymbol{\alpha} = \mathbf{0} \tag{2.9}$$

where $S^1_{0,27}(\boldsymbol{T}_{it})$ denotes a linear spline in the full time series of temperature, assuming single sine interpolation within each day, with knots at 0°C and 27°C.[6] The base specification is intended to represent the case where the effect of measurement error in control variables leaves no useful signal in these variables.

The first specification with controls introduces a quadratic in total season precipitation:

$$g(\boldsymbol{P}_{it}) = \beta_{p1}P_{it} + \beta_{p2}P^2_{it} \tag{2.10}$$

Here, P_{it} denotes total season precipitation. This is the specification that is used in both Schlenker and Roberts (2009) and Schauberger et al. (2017).

[6]The apparent turning point for the relationship between temperature and yields in the ISIMIP data tends to be 27°C; for AgMERRA, 25°C appears to be the optimal turning point. Note that these datasets have lower spatial aggregation than the dataset used in Schlenker and Roberts (2009), which finds a turning point of 29°C, so there is more spatial averaging of temperature extremes.

The next specification replaces the precipitation quadratic with a flexible linear spline in daily precipitation:

$$g(\boldsymbol{P}_{it}) = S^1_{0,1,5,10,15,20}(\boldsymbol{P}_{it}) \tag{2.11}$$

where $S^1_{0,1,5,10,15,20}(\boldsymbol{P}_{it})$ denotes a linear spline in daily precipitation with knots at 1mm, 5mm, 10mm, 15mm, and 20mm.

Next, I investigate the effect of additional controls by adding flexible linear splines in each of solar radiation, relative humidity, and windspeed:

$$\boldsymbol{z_{it}\alpha} = h_1(\boldsymbol{R}_{it}) + h_2(\boldsymbol{H}_{it}) + h_3(\boldsymbol{W}_{it}) \tag{2.12}$$

$$h_1(\boldsymbol{R}_{it}) = S^1_{0,2,4,8,12,16,20,24,28,32,36}(\boldsymbol{R}_{it}) \tag{2.13}$$

$$h_2(\boldsymbol{H}_{it}) = S^1_{0,21,39,60,81}(\boldsymbol{H}_{it}) \tag{2.14}$$

$$h_3(\boldsymbol{W}_{it}) = S^1_{0,2,4,6,8}(\boldsymbol{W}_{it}) \tag{2.15}$$

where the spline functions are defined similarly to that for precipitation. To ensure that the simple specification of temperature does not obscure important features of the temperature responses, as a robustness check I implement a flexible specification that replaces (2.7) with:

$$f(\boldsymbol{T}_{it}) = S^1_{-5,0,5,10,15,17,19,21,23,25,27,29,31,33,35}(\boldsymbol{T}_{it}) \tag{2.16}$$

which is a linear spline in the full time series of temperature with knots at the values in the subscript.

In all specifications that use the ISIMIP data, I pool the observations across all five climate models, for each crop model, and I allow the county fixed effects and state-specific quadratic trends, but not weather responses, to vary across climate models.

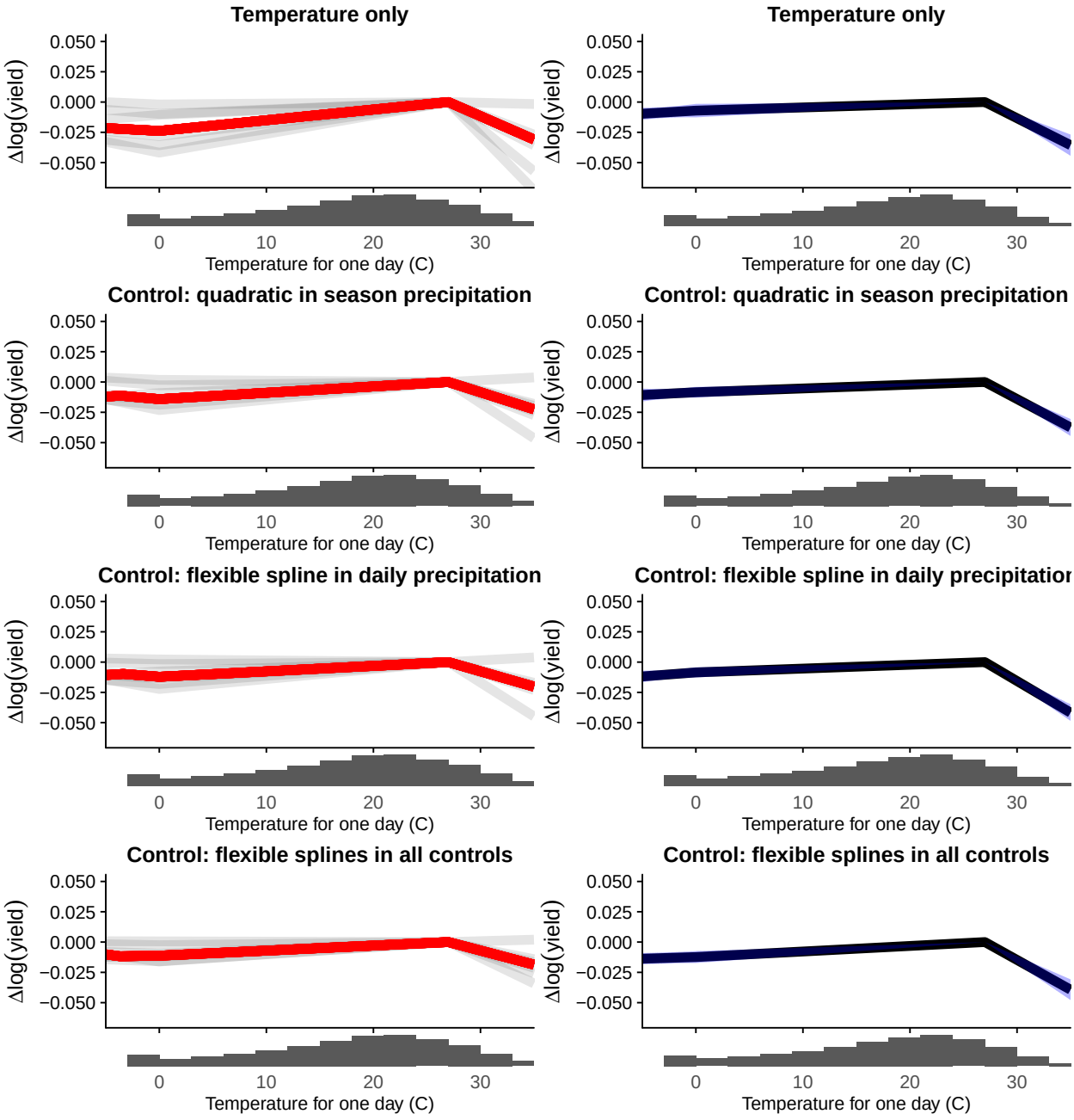
Examining residual correlations

As alluded to in the Theory section, Garber and Klepper (1980) show that the bias in the coefficient on a variable of interest caused by mis-measurement of a control variable depends on the correlation between the two variables, after partialling out the effects of all other included variables. The true bias depends on the correlation between the true, rather than the observed, variables, but we can use the observed variables as proxies, so long as the signal contained in the mis-measured variable is large enough. If the measurement error is classical, the correlation between the observed variables should be lower than the correlation between the true variables, so using observed variables should be a conservative indicator of potential bias.

2.5 Results

Figure 2.1 plots the results for both the process (column 1) and empirical (column 2) models using the temperature specification in Equation (2.7). The rows of the figure use increasingly

Figure 2.1: Process model and empirical model temperature response functions



This figure plots Equation (2.7) *for a single day of temperature. Each plot is vertically centered so that change in log(yield) takes a value of zero when temperature is 27° C. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of temperature. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of the full time series of temperature using single sine interpolation.*

detailed controls: the first row only containing temperature variables, the second row using a quadratic in season-level precipitation (Equation (2.10)), the third using a flexible linear spline in daily precipitation (Equation (2.11)), and the fourth using flexible linear splines in each of precipitation, solar radiation, relative humidity, and wind speed (Equations (2.11), (2.13), (2.14), and (2.15)). I plot the results for the individual process models in light gray and the median value for each value of temperature in red; all the process model values are point estimates. I plot the results for the empirical models with 95% confidence bands that are calculated assuming error clustering by state and year.

Firstly, I find that the median point estimate within the process models for the impact of extreme temperature, when including no control variables, is remarkably similar to that for the empirical model. Two process models show much larger temperature effects and one process model shows almost no temperature effect. The similarity between the median process model function and the empirical function is consistent with measurement error in temperature being sufficiently small such that the temperature response does not substantially attenuate, if the median process model response is correct.

Introducing the precipitation quadratic reduces the median extreme temperature effect in the process models to 73% of the slope in the no-controls regression, while it *increases* the slope on extreme temperature in the empirical model by around 5%. Using the flexible precipitation specification further reduces the extreme temperature effect in the process models to 66% of the slope in the no-controls regression,[7] while slightly increasing the slope further in the empirical model. Introducing the other control variables further reduces the extreme temperature effect in the process models to 60% of the slope in the no-controls regression. If the median process model response is correct, these results are consistent with substantial measurement error in the control variables causing bias in the coefficient on extreme temperature in the empirical model.

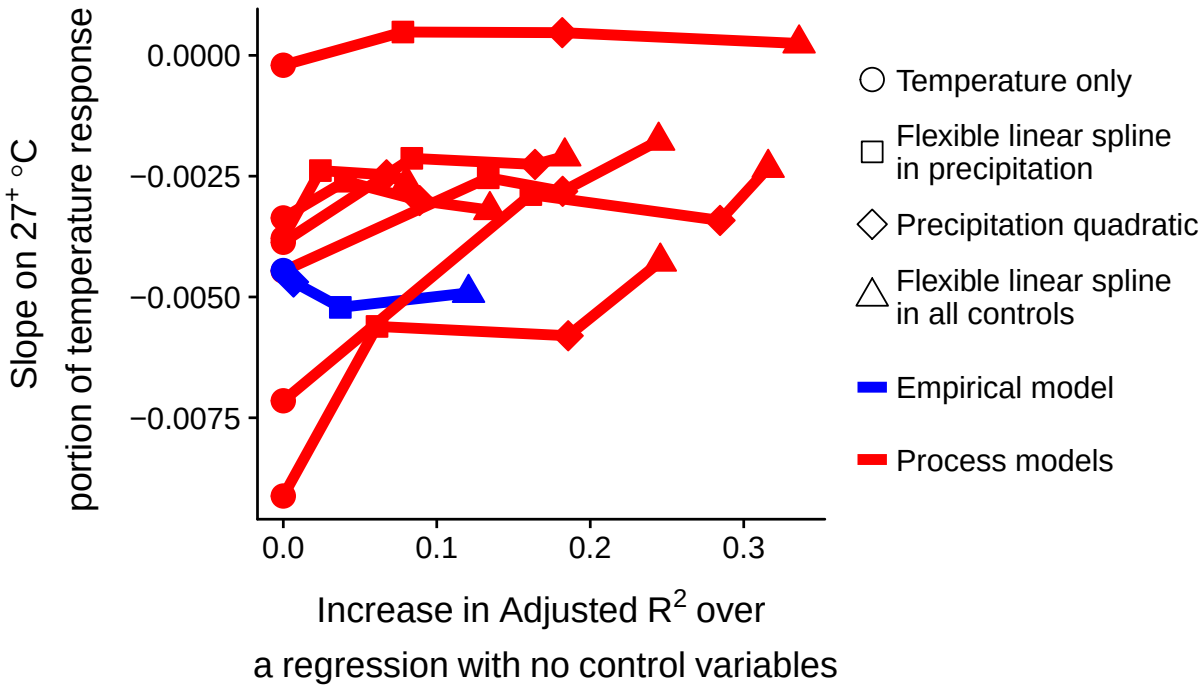
Focusing on the median obscures some substantial heterogeneity amongst the process models. Figure 2.2 plots the extreme temperature slope against the increase in the R^2 over a regression with no control variables. The values within each of the seven process models as well as the empirical model are joined by lines. This plot is intended to communicate the degree to which each set of controls both explains variation in yields and shifts the extreme temperature response. If controlling for more variables tends to reduce the observed impact of extreme temperature, then the plotted lines should be upward sloping.

The figure shows three stark features: all seven process models' extreme temperature effects reduce when including controls, the models with the largest temperature effects have the largest reductions, and all seven process models' extreme temperature effects are smaller than the empirical model's, after including all controls.

Figure 2.3 shows the precipitation response functions for the three specifications in Figure

[7]Note that the season-level quadratic actually explains more variation in annual yields than the daily model. However, the using the daily precipitation variation might control for more of the portion of precipitation that varies with extreme temperature.

Figure 2.2: Extreme temperature slopes against increase in R^2, by model

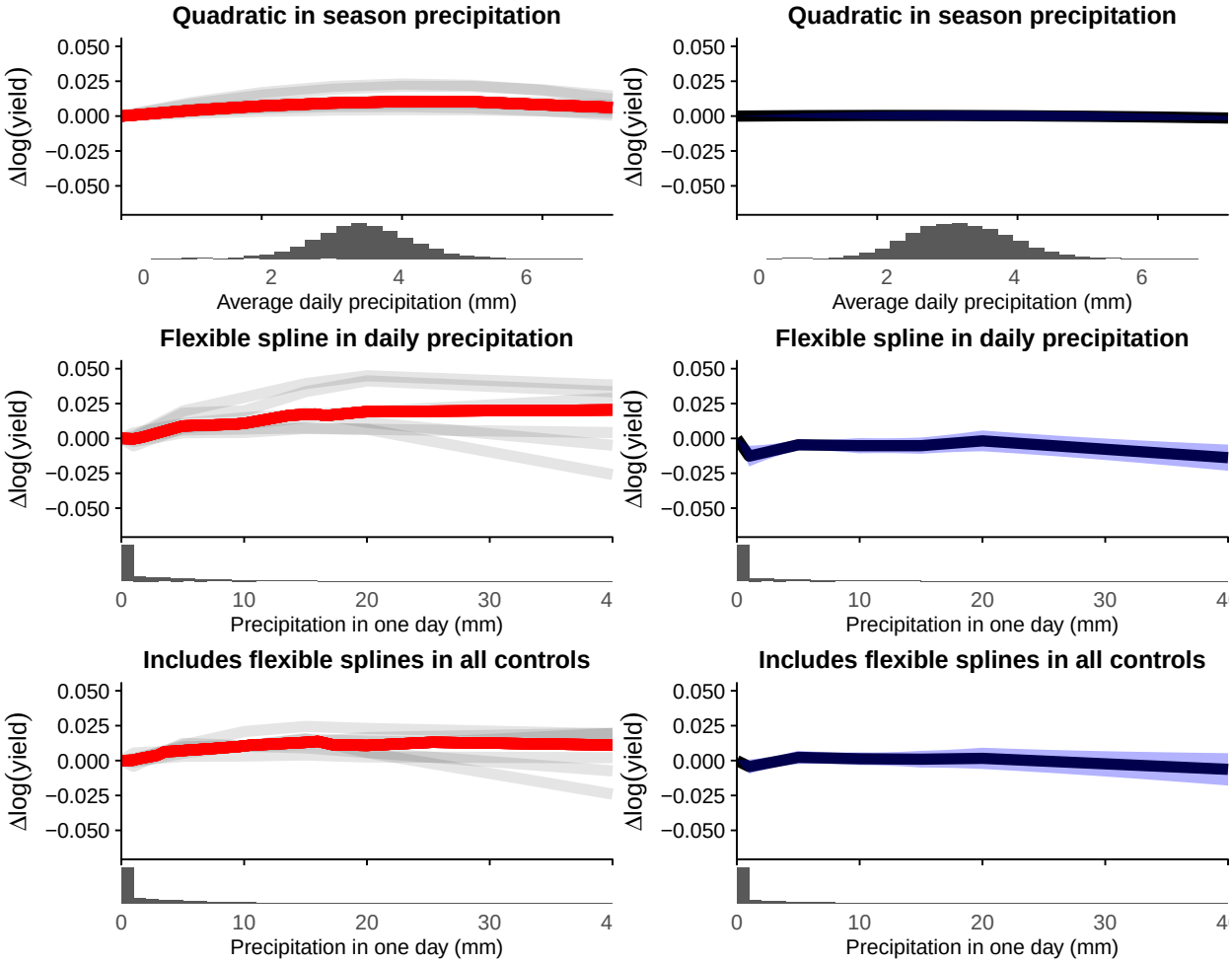


This figure plots the slopes on the portion of the temperature response function above 27° C against the increase in adjusted R^2 over a regression including temperature variables only, for each of the specifications in Figure 2.1, and each of the seven process models as well as the empirical model. The figure links the four specifications for each of the eight models with a line.

2.1 that include precipitation. The season-level quadratic functions in the process models are all inverted-U-shaped, though quite flat where most of the distribution of rainfall lies. The empirical precipitation response function, however, is a very tight zero, consistent with what we would expect when using a highly mis-measured regressor. In the model with flexible linear splines in daily precipitation and no other control variables, I find that the process models generally have an inverted-U relationship between daily precipitation and yield. Oddly, the empirical response function features a strange large and negative effect of small amounts of precipitation. Absent this feature, the precipitation response function appears to be qualitatively similar to the median response in the process models. Introducing other controls, including solar radiation, substantially reduces the size of the negative impact of small amounts of precipitation in the empirical model, indicating that it was likely omitted solar radiation that caused the strange negative effect of small amounts of precipitation.[8] However, introducing the extra controls also reduces the estimated positive effects

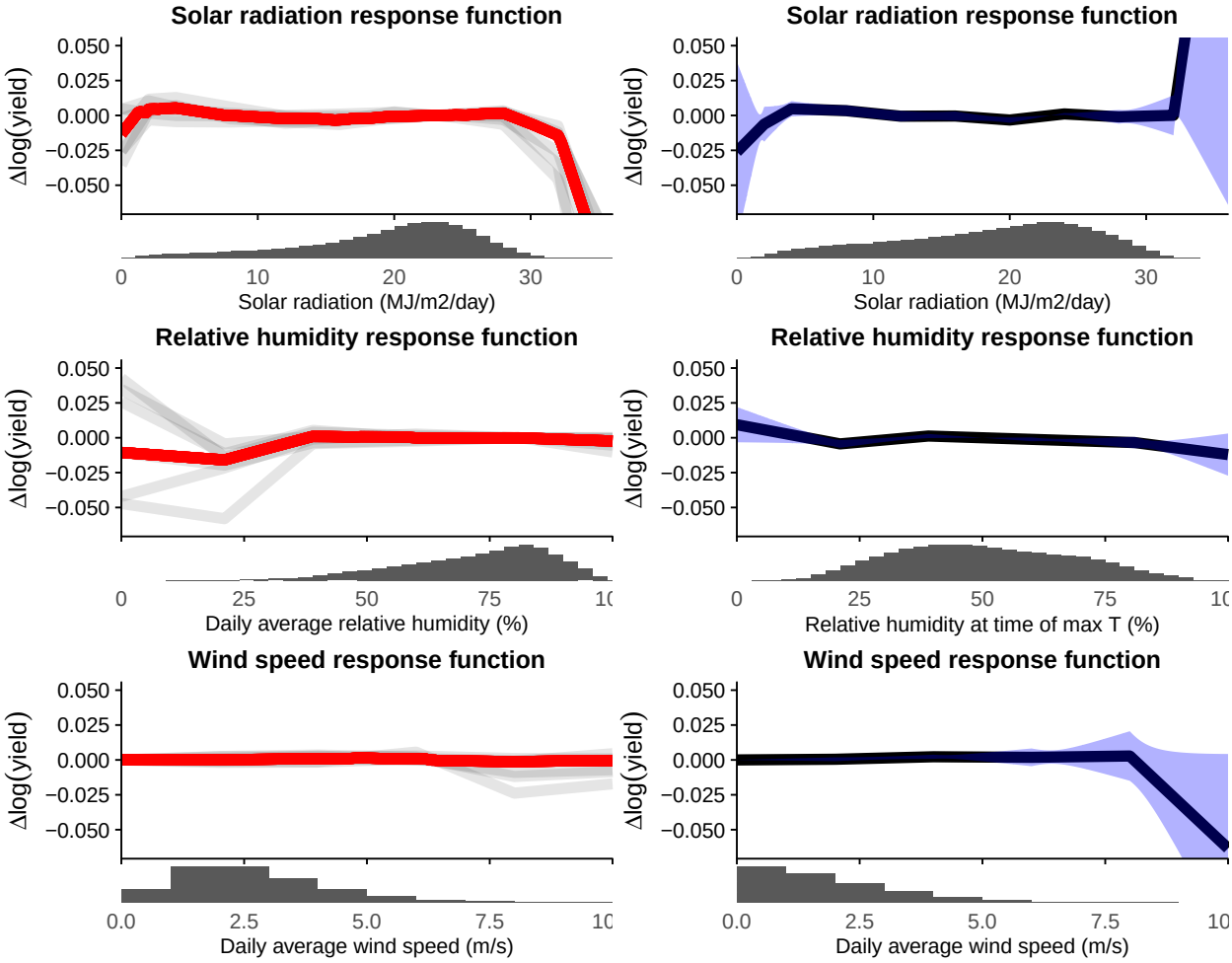
[8]Very low levels of precipitation should have little impact on yields compared to no precipitation, but cloudy days also have less sunlight, so solar radiation is likely negatively correlated with the low-precipitation

Figure 2.3: Process model and empirical model precipitation response functions



This figure plots Equations (2.10) *for average daily precipitation and* (2.11) *for a single day of precipitation. Each plot is vertically centered so that change in log(yield) takes a value of zero when precipitation is 0mm. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of precipitation. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below the plots in the first row is a histogram of average daily precipitation. Below the plots in the second and thirds rows are histograms of daily precipitation.*

Figure 2.4: Process model and empirical model flexible response functions to other control variables



This figure plots Equations (2.13), (2.14), *and* (2.15) *for a single day. Each plot is vertically centered so that change in log(yield) takes a value of zero when solar radiation is 23MJ/m^2/day, daily average relative humidity in the process models is 80%, relative humidity at the time of maximum temperature in the empirical model is 50%, and wind speed is 0m/s. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of the independent variable. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles indicate which control variable is being plotted. Below each plot is a histogram of the daily values of each variable.*

of precipitation, leaving the entire function statistically indistinguishable from zero. The median crop model precipitation response does not change substantially when introducing other controls.

---

regressors.

Figure 2.4 plots the responses for the control variables other than precipitation. Despite a large amount of variation in solar radiation in both the process and the empirical models, neither model seems to recover the positive effect across the spectrum that we expect.

We expect relative humidity to have an inverted-U-shaped relationship with yields, with low humidity increasing water demand, and high humidity potentially increasing disease susceptibility. In the empirical model, humidity appears to have a small inverted-U-shaped response in the main part of the support of the independent variable, while the process models show mostly flat responses.

Wind speed appears to have very little effect in the main part of its support in both the process and empirical models.

To ensure the simple base specification of temperature is not obscuring features of the underlying response functions, Figure 2.5 replaces the temperature specification in Figure 2.1 with a flexible linear spline. The functions look generally similar to those in the simpler model, with temperatures starting to worsen yields close to 27°C. There is, however, substantial heterogeneity and "spikiness" in the process models' responses. The empirical temperature response closely mirrors the simpler specification's.

The lack of smoothness in the flexible temperature responses highlights a problem in the emulation of the process models. While the process model emulations do not suffer from measurement error, they *do* suffer from specification error.[9] It is unclear exactly how incorrect specification would affect the bottom-line conclusions of this chapter. To remedy this problem, a future iteration of this project will attempt to identify the complicated relationships between annual simulated yields and the many possible input variables using LASSO regression, a machine learning technique that will help to choose which of the many possible specifications best represents each process crop model.[10]

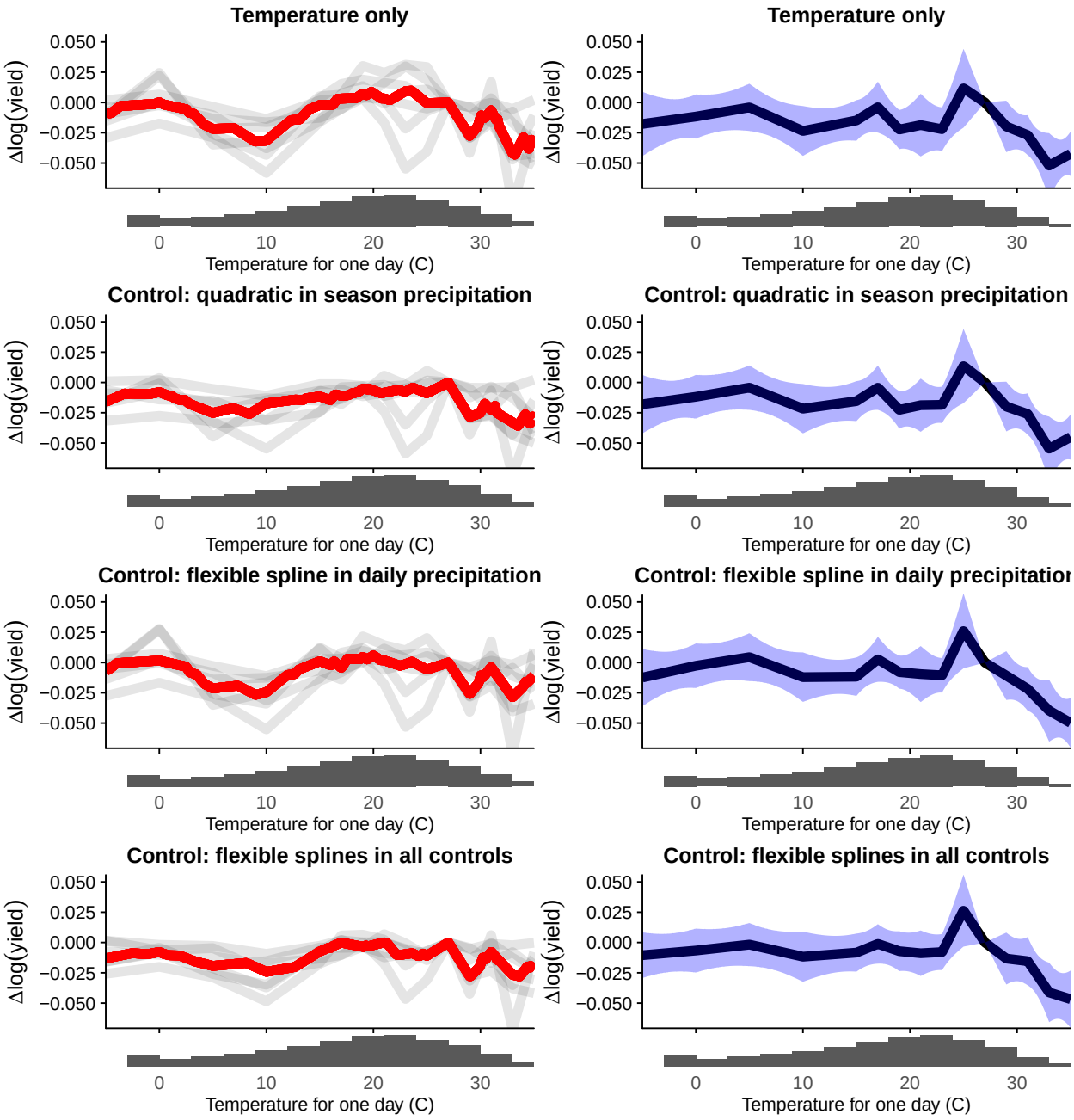
Examining residual correlations

Table 2.1 presents the correlations between each of the temperature variables in Equation (2.7) and precipitation, after partialling out the effects of all fixed effects, trends, and the other temperature variables. For simplicity, I omit the square of the precipitation variable when calculating these correlations. The table shows a strikingly large and negative residual correlation between total season precipitation and the number of degree days in a season above 27°C. Because the effect of precipitation is positive, Equation (2.4) states that the bias in the coefficient on the temperature regressor is negative. This diagnostic is unable

[9]Throughout, I use "specification error" to refer to omitted variable bias that arises from correlations between included regressors and omitted nonlinearities and seasonal heterogeneity in response functions. Correlation between the extreme temperature regressor and the cube of season-level precipitation would be a possible source of "specification error", for example.

[10]There are six weather inputs for 184 days per growing season, with non-linear transformations, and interactions.

Figure 2.5: Process model and empirical model flexible temperature response functions



This figure plots Equation (2.16) *for a single day of temperature. Each plot is vertically centered so that change in log(yield) takes a value of zero when temperature is 27° C. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of temperature. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of the full time series of temperature using single sine interpolation.*

Table 2.1: Residual correlations between non-linear temperature and precipitation

	Temperature regressors ($S^1_{0,27}$)		
	$T < 0$	$0 \leq T < 27$	$T \geq 27$
Residual correlation with season precipitation	-0.041	0.118^{*}	-0.526^{***}
	(0.063)	(0.064)	(0.067)

Following Garber and Klepper (1980), this table presents residual correlations between each temperature regressor in the linear spline with knots at 0°C and 27°C and season level precipitation, after controlling for state-specific quadratic trends, county fixed effects, and the other temperature regressors. I calculate the standard errors for these correlations assuming error clustering by state and year. For simplicity, I omit the square of the precipitation variable when calculating these correlations. If the effect of precipitation is positive, and only precipitation is measured with error, the bias in the temperature coefficient is the same sign as the reported correlation. The magnitude of the bias depends on the signal-to-total variance ratio for the residualized precipitation variable, the size of the true precipitation effect, and the ratio of the standard deviations of the residualized precipitation variable and the residualized temperature variable. "*", "**", and "***" denote statistical significance at the 10%, 5%, and 1% levels respectively.

to give any guidance as to how large the bias might be, as that requires knowledge of the magnitude of the measurement error in precipitation.

2.6 Possible solutions

This section briefly outlines some suggested solutions from the literature, implements the proposed long differences solution, and outlines my proposed instrumental variable solution.

Test for measurement error using Griliches and Hausman (1986)

As implemented in Burke and Emerick (2016), Griliches and Hausman (1986) suggest comparing the coefficients from a random effects and a fixed effects model to indicate if measurement error is causing problems in the latter. However, while the random effects model will suffer less from the measurement error problem, it will suffer more from the omitted variables bias problem. In the random effects model, it's possible that the increase in omitted variables bias and the decrease in measurement error bias could offset each other and result a coefficient similar to the fixed effects model. Thus, it's unclear if this test is a useful diagnostic in this context.

Using reverse regressions

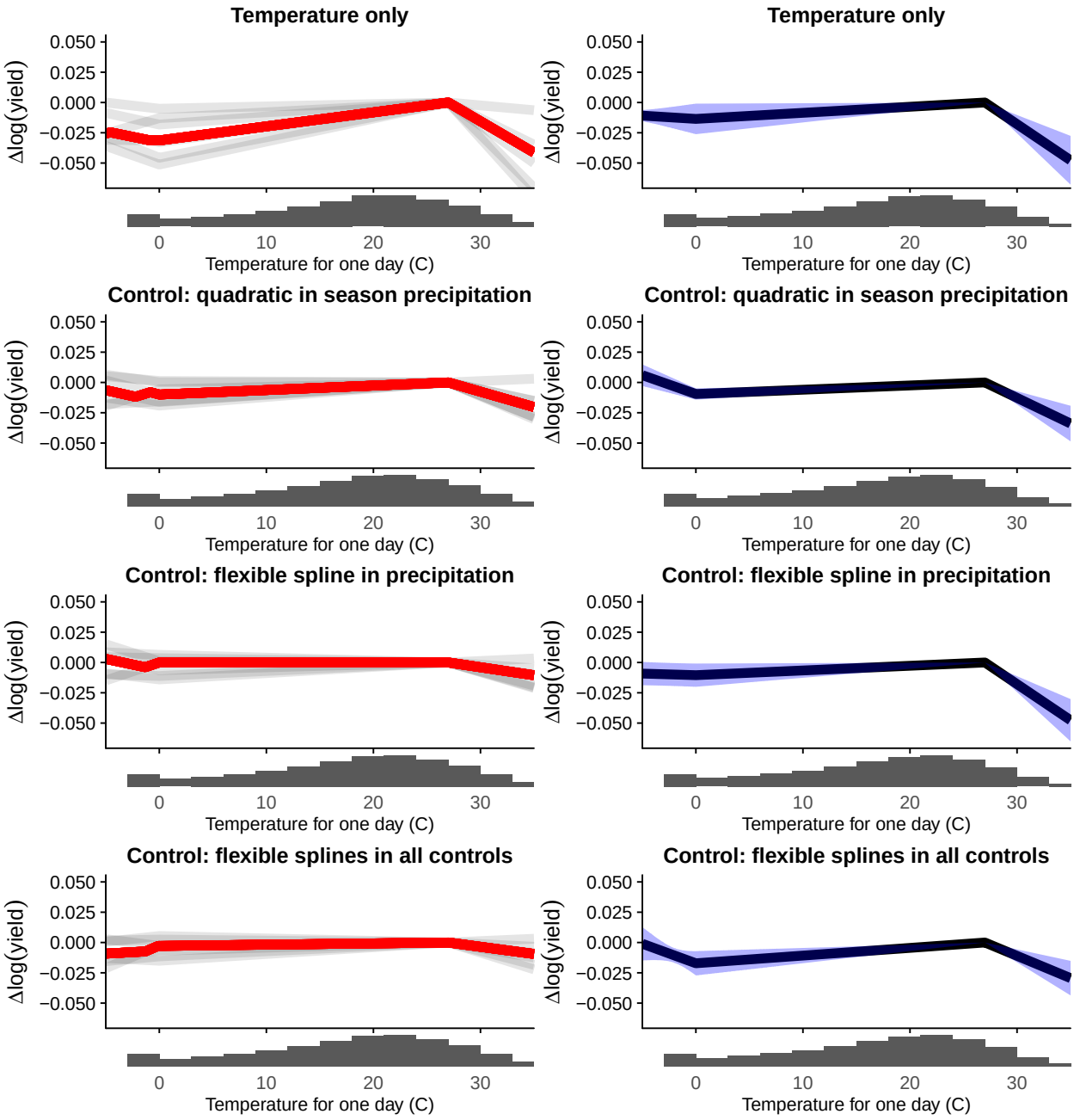
Following Klepper and Leamer (1984), I experimented using a reverse regression to bound the true value of the temperature response. The specification used Equations (2.7) and (2.10), and reversed the two precipitation regressors. The implied coefficients on the extreme temperature variable in the two reverse regressions were implausibly large and positive, suggesting that this method is unlikely to be a viable solution in this context.

Long differences

Burke and Emerick (2016) use a long differences model to detect the fingerprint of adaptation to climate change in the context of maize in the USA. They posit that farmers may react to longer run changes in climate differently to short-run changes and thereby reduce the impact of extreme temperature on their yields. However, this method could also partially remedy the problem of measurement error in precipitation for two reasons. Firstly, the long differences model uses fewer time effects, increasing the proportion of signal used by the estimator. Secondly, the longer run temperature change *may* reduce the problematic correlation between extreme temperature and precipitation. However, it is then unclear how to distinguish the effect of farmer adaptation from the effect of reductions in measurement error in precipitation, both of which should reduce the impact of extreme temperatures. If measurement error in precipitation is indeed a problem here, any apparent adaptation observed by the long differences model may be spurious.

This section replicates the analysis in this chapter thus far, using long differences. Following Burke and Emerick (2016), I sum the observations within each unit for each of the first five and the last five years in the respective samples. As the samples now only have two

Figure 2.6: Process model and empirical model temperature response functions in the long differences model



This figure plots Equation (2.7) for a single day of temperature. Each plot is vertically centered so that change in log(yield) takes a value of zero when temperature is 27° C. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of temperature. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of the full time series of temperature using single sine interpolation.

time periods, I omit the state-specific trends and include a sample-level time fixed effect in the empirical model, and a climate-model-specific time fixed effect in the process models.

Figure 2.6 replicates the specifications in Figure 2.1, using long differences. As expected, the empirical response function is now more sensitive to the inclusion of control variables, consistent with the long differences model lessening the problem of measurement error in precipitation. In the empirical model, including the precipitation quadratic decreases the temperature response to around 71% of that for the regression with no controls. However, including the flexible spline in daily precipitation increases the temperature response back to 100% of that for the no-controls specification. Including flexible splines in all control variables decreases the temperature response to 62% of that for the regression with no controls.

While the results for the empirical model are encouraging, what is discouraging is that the sensitivity to the inclusion of controls in the process models has also increased. Introducing the precipitation quadratic, the flexible spline in daily precipitation, and flexible splines in all control variables respectively reduces the median extreme temperature response to 49%, 26%, and 23% of that for the regression with no controls.

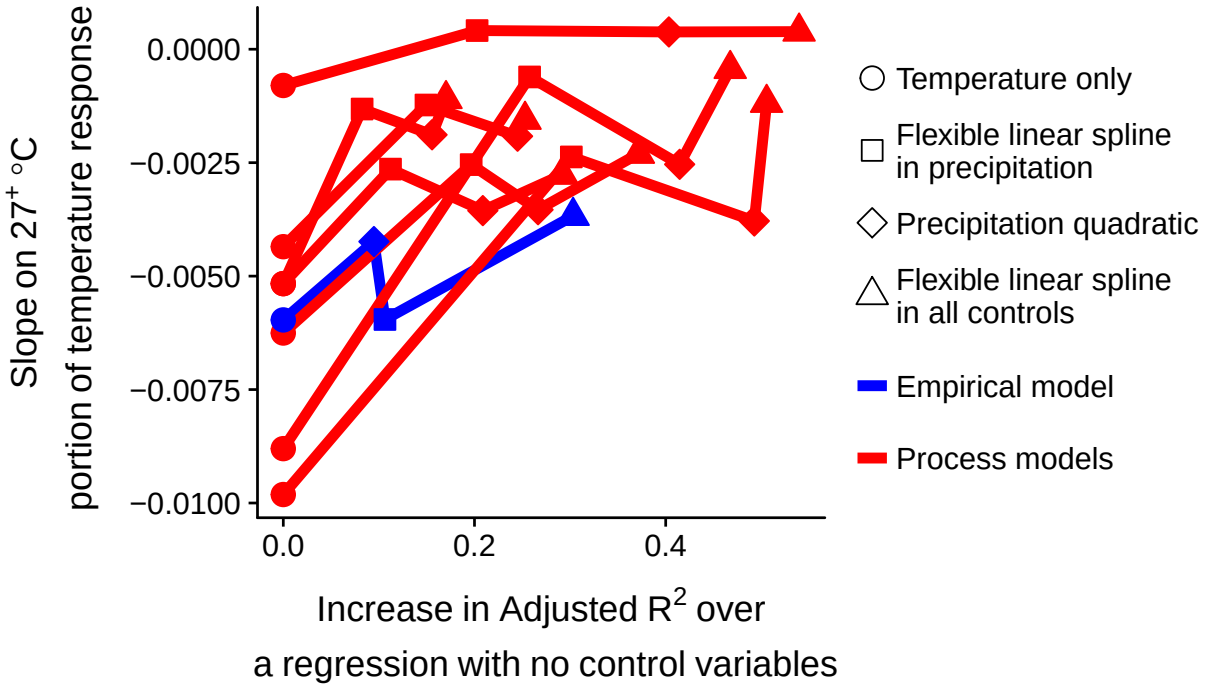
While the results appear puzzling, it is possible that the process models are suffering from omitted variables bias (or specification error), and the correlation between temperature and the omitted variables changes when using the different time scales for estimation. It's not clear which direction this bias might go. For this reason, it's important that a future iteration of this project can more confidently determine the underlying functions in the process models before drawing hard conclusions.

Figure 2.7 summarizes the results of the previous figure by plotting the slopes on the extreme temperature portion of the response function against the gain in R^2 over a regression with no controls. The most notable difference, compared with the equivalent plot that used the fixed effects specifications, Figure 2.2, is that the process models with the largest temperature impact in the no-controls specification have such large reductions that they become the second and third *smallest* temperature impacts in the specification with flexible splines in all control variables. As in the previous plot, introducing controls reduces the temperature response in all the process models.

Figure 2.8 shows the precipitation response functions. As expected, the empirical response for the season-level quadratic model is now statistically significant, and is close to that for the median response from the process models. However, the flexible spline specification still shows large negative effects of small amounts of precipitation on yields. In the previous section, I suggested that the omission of solar radiation could have caused this anomaly; however, in the long differences specification, introducing solar radiation does not substantially reduce this negative effect. Examining Figure 2.9 shows that the empirical solar radiation response function slopes downward for much of the support. The downward slope suggests that mis-measurement or mis-specification in other variables is causing bias in the response to solar radiation, causing some new concern for the robustness of the empirical model.

Figure A2 plots the responses using the flexible temperature specification. Both the

Figure 2.7: Extreme temperature slopes against increase in R^2, by model in the long differences model



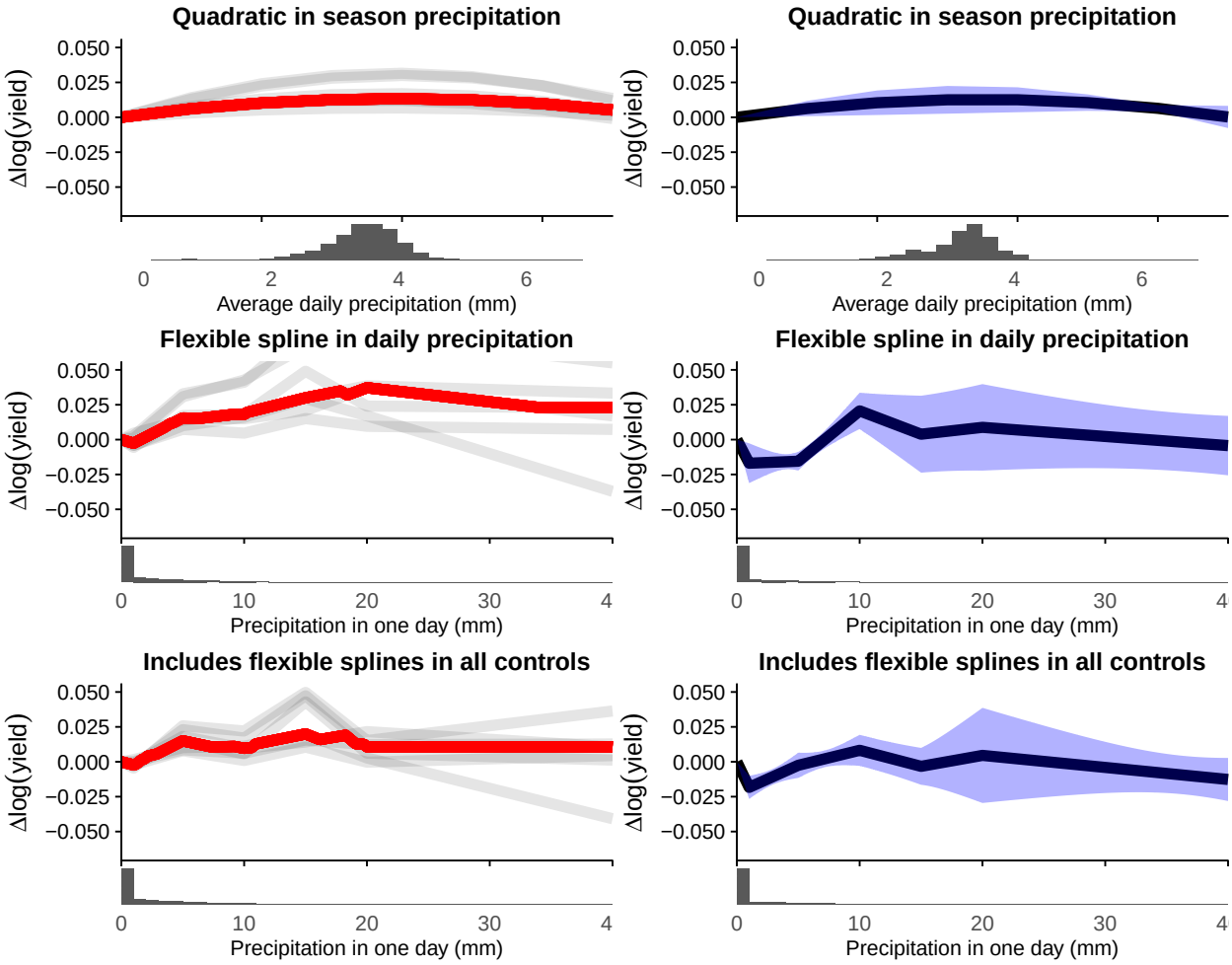
This figure plots the slopes on the portion of the temperature response function above 27°C against the increase in adjusted R^2 over a regression including temperature variables only, for each of the specifications in Figure 2.1, and each of the seven process models as well as the empirical model. The figure links the four specifications for each of the eight models with a line.

responses in the process and empirical models are quite noisy, so it is difficult to draw conclusions from this figure.

The final piece to replicate for the long differences model is to examine the residual correlations between the temperature regressors and precipitation. Indeed, I do find that the negative correlation between the extreme temperature regressor and precipitation is less than that in the fixed effects model. However, it remains large and statistically significant at -0.45. Unfortunately, the long differences variation also creates significant correlations between precipitation and the cold and moderate temperature regressors of 0.309 and -0.221 respectively, causing concern about possible bias in the coefficients on those variables.

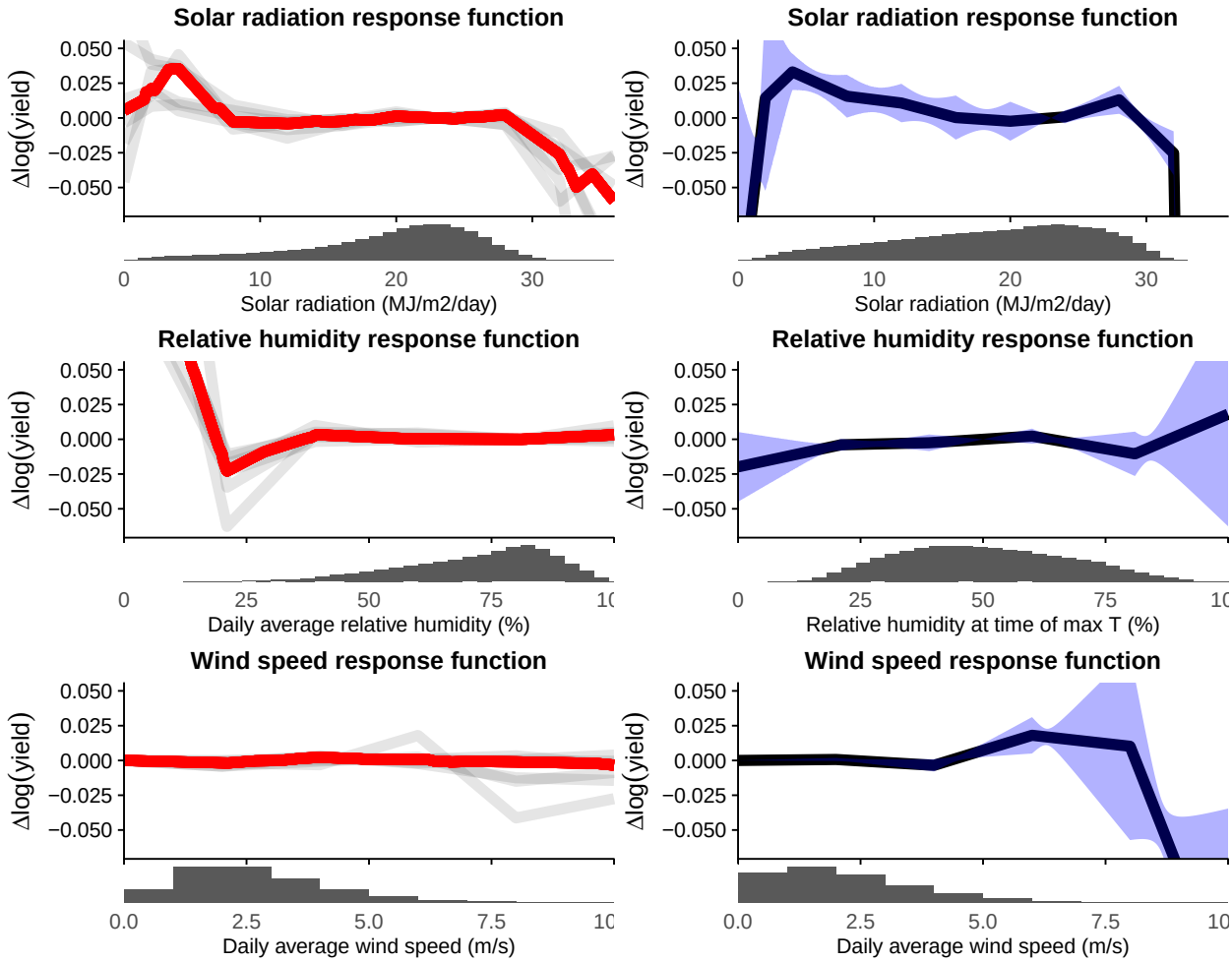
To summarize the results from the long differences specifications, I find that the empirical temperature response functions are more sensitive to the inclusion of controls, and that the empirical control variables explain more variation in yields, when compared with the fixed effects model. However, I also find that the temperature response functions for the process models are also more sensitive to the inclusion of controls, potentially indicating

Figure 2.8: Process model and empirical model precipitation response functions in the long differences model



This figure plots Equations (2.10) *and* (2.11) *for a single day of precipitation. Each plot is vertically centered so that change in log(yield) takes a value of zero when precipitation is 0mm. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of precipitation. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of daily precipitation.*

Figure 2.9: Process model and empirical model flexible response functions to other control variables in the long differences model



This figure plots Equations (2.13), (2.14), *and* (2.15) *for a single day. Each plot is vertically centered so that change in log(yield) takes a value of zero when solar radiation is 23MJ/m^2/day, daily average relative humidity in the process models is 80%, relative humidity at the time of maximum temperature in the empirical model is 50%, and wind speed is 0m/s. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of the independent variable. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles indicate which control variable is being plotted. Below each plot is a histogram of the daily values of each variable.*

Table 2.2: Residual correlations between non-linear temperature and precipitation in the long differences model

	Temperature regressors ($S^1_{0,27}$)		
	$T < 0$	$0 \leq T < 27$	$T \geq 27$
Residual correlation with season precipitation	0.309^{***}	-0.221^{***}	-0.452^{***}
	(0.082)	(0.086)	(0.1)

Following Garber and Klepper (1980), this table presents residual correlations between each temperature regressor in the linear spline with knots at 0°C and 27°C and season level precipitation, after controlling for state-specific quadratic trends, county fixed effects, and the other temperature regressors. I calculate the standard errors for these correlations assuming error clustering by state and year. For simplicity, I omit the square of the precipitation variable when calculating these correlations. If the effect of precipitation is positive, and only precipitation is measured with error, the bias in the temperature coefficient is the same sign as the reported correlation. The magnitude of the bias depends on the signal-to-total variance ratio for the residualized precipitation variable, the size of the true precipitation effect, and the ratio of the standard deviations of the residualized precipitation variable and the residualized temperature variable. "*", "**", and "***" denote statistical significance at the 10%, 5%, and 1% levels respectively.

that those models suffer from specification error. I find that the empirical precipitation response function when using the quadratic specification is reasonable, and similar to that for the process models; however, when using the flexible linear spline I still find large negative effects of small amounts of precipitation, as with the fixed effects model. I also find that the empirical solar radiation response function slopes downward for much of the support, suggesting that there are still measurement problems in the long differences specification for the empirical model.

Using interpolated weather as an instrument

The most optimistic path forward, for contexts in which researchers expect precipitation impacts to be large, and precipitation monitors are somewhat dense, is computing two interpolations of each grid cell with different sets of monitors for each grid cell-day, then instrument one interpolation with the other. However, this relies on correct spatial aggregation within each unit and time period as incorrect aggregation leaves relevant variables in the error term, and the weather at the incorrectly omitted grid cell is likely to be more correlated with one of the sets of monitors more than the other. A future iteration of this project will implement this solution.

2.7 Conclusion

This chapter gave several pieces of evidence that suggest that standard empirical maize yield models overstate the impact of extreme temperature due to mis-measurement of control vari-

ables. Specifically, the chapter finds that temperature response functions in maize process models have much higher sensitivities to the inclusion of control variables than their empirical counterpart. Importantly, the chapter also documents that extreme temperature and precipitation have a large negative correlation after removing the effects of other variables and fixed effects, consistent with what would be required for mis-measurement of precipitation to cause bias away from zero in effect of extreme temperature.

But does this finding have implications for the wider field of weather impacts estimation for sectors other than agriculture? I argue that the answer is yes, for any sector for which we believe that variables other than temperature might have an impact, and those variables are either omitted or measured with error. Temperature and humidity, for example, interact in the relationship between weather and mortality; precipitation also reduces air pollution, which contributes to mortality. Precipitation is important for agricultural mechanisms in the relationships between weather and conflict or migration. Outdoor labor supply and productivity depends on both temperature and precipitation. Future research could re-examine each of these contexts with an eye to determining any possible biases from omission or mis-measurement of control variables and, if the biases might be substantial, attempt to correct for the problem using an instrumental variables estimator based on multiple spatial interpolations.

Importantly, the finding of this chapter does not necessarily imply that we should expect to find that prior studies have overstated the effects of temperature; the opposite is possible.

Chapter 3

Empirical estimation of the impact of weather on dairy production

3.1 Introduction

Agricultural production depends heavily on the weather, causing widespread and early concern about the effects of global climate change on this sector. In the economics literature, seminal papers have focused on the effects of average temperature on agricultural land values (Mendelsohn, Nordhaus, and Shaw 1994; Schlenker, Hanemann, and Fisher 2005), the effects of contemporaneous shocks to temperature on agricultural profits (Deschênes and Greenstone 2007; Fisher et al. 2012; Deschênes and Greenstone 2012), and the nonlinear effects of contemporaneous temperature shocks on several crop yields (Schlenker and Roberts 2009) (for other work on this topic, see Auffhammer and Schlenker (2014) for a review). More recent research has combined data from agronomic models with global agricultural trade models to suggest that reallocation of crops can moderate projected negative impacts of climate change by around a third (Costinot, Donaldson, and Smith 2014).

However, while crops provide the majority of the supply of most nutritional variables for most places, livestock systems comprise a very large portion of the agricultural system and a substantial portion of nutrition. For example, livestock systems account for most of the land used in agriculture, with the vast majority of this land in pasture, and livestock contributes around 30% of the global protein supply, with this proportion higher in the developed world and projected to increase in the developing world. To date, there has been little work that has examined the effects of climate change on any livestock systems. In the livestock context, changing climates could affect each of feed production, the level of production of final output per unit of feed, and the quality of the final product, potentially resulting in a triple whammy for the food system.

This chapter extends the climate and agriculture literature in two ways. Firstly, using data from New Zealand, the largest dairy exporter and a major producer, it examines the

effect of climate change on the largest animal sector, as measured by the contribution to both the global protein and calorie supply, dairy, using an empirical panel fixed effects approach. Secondly, it allows for heterogeneity in response functions by time-of-year and examines the impact of this modeling choice on bottom-line conclusions about the costs and benefits of future climate change. In addition to the contributions to the literature on climate and agriculture, this chapter also provides calculations of the relative contributions of the major land uses to the global food supply, including those from pasture.

Because livestock is a weather sensitive industry, with weather sensitive feed inputs that may be transported some distance before being consumed by the animal, identification of climate change impacts using weather impacts is complicated by the need to account for upstream impacts in the feed industry. New Zealand is an attractive setting to study the potential impact of climate change on a livestock sector, as it is both a large producer and an industry that primarily uses local feed. Anecdotally, over the time period I study, only around 10% of the feed used is from a source other than direct intake from pasture, and a large proportion of the supplemental feed either comes from on farm or the nearby area. This suggests that I'm able to identify the impact of future climate change using local weather shocks, conditional on the assumption that the historical relationship between weather and output continues to hold (Hsiang 2016).

When restricting the temperature response function to be the same throughout the year, I find large and negative impacts of moderate to high temperatures above 19°C, with a statistically insignificant response to temperature below 19°C. However, when allowing for seasonality in the response function, I still find large and negative impacts of moderate to high temperatures in summer months but the model also shows large and positive impacts of moderate temperatures during winter months. When projecting these response functions forward, using the output from a climate model, I find that the discounted present value of the projected change in revenue is -US$8.4 billion when using the restricted model, and +US$2.1 billion when using the flexible model. The total farm gate value of annual New Zealand dairy production is currently approximately US$8 billion.

Existing evidence on the effect of temperature in dairy systems has shown reductions in dry matter intake with moderate to extreme temperatures, causing reductions in milk output (West 2003). Pasture growth also responds negatively to high temperatures (Cros et al. 2003). Work has also demonstrated direct effects on production of heat stress when controlling for feed intake, as well as suggestive evidence that protein percentage slightly decreases under heat stress (Rhoads et al. 2009). Studies have estimated the production effect of cattle heat stress using a stochastic frontier approach (Mukherjee, Bravo-Ureta, and De Vries 2013; Qi, Bravo-Ureta, and Cabrera 2015). One work combined process and climate models to estimate the impact of climate change on dairy production in the Australian context (Hanslow et al. 2014).

Several studies have reported functions that allow for different responses to weather at different times of the year (Welch et al. 2010; Cooper, Nam Tran, and Wallander 2017; Schlenker and Roberts 2009). I extend the ideas in these studies by examining the impact

of this modeling choice on climate change projections. Results from financial econometrics suggest that bias can result by restricting responses to be the same throughout the season (Andreou, Ghysels, and Kourtellos 2010).

This chapter proceeds as follows. Firstly, the following section describes where both the dairy industry and pasture-based livestock systems sit in relation to the larger agricultural system. Next, Section 3.3 describes the sources of my dairy production and weather data, Section 3.4 describes the theoretical concepts that inform the analysis, Section 3.5 outlines the econometric specifications I use, Section 3.6 reports my results and projections of the consequences of climate change in this context, and Section 3.6 concludes.

3.2 The importance of dairy and pasture

Since staple crops dominate the climate and agriculture literature, it is useful to consider where each of dairy and pasture stand in relation to the rest of the food system. Table 3.1 ranks FAO (2014) food balance data for the proportion of world protein coming from different food groups. The FAO food balance data aims to measure the total food available for human consumption by type and country. I present global aggregates in this chapter.

Despite their relatively low ratio of protein to other calories, the prevalence of wheat and rice in the global food system makes them the largest contributors to protein consumed, with 19.7% and 12.7% of the world totals, respectively. Dairy is the largest contributor to global protein consumed amongst both animal products and high protein foods (i.e. including beans and pulses) generally. Table 3.2 provides the same ranking for global calorie contributions. Again, wheat and rice are the dominant calorie sources with 18.3% and 19% respectively, with dairy again contributing the largest proportion amongst animal products with 5.9%.

Table 3.1: Foods ranked by contribution to world protein

FAO Food Balance Item	World Protein Percentage
Wheat and products	19.7%
Rice (Milled Equivalent)	12.7%
Dairy	10.3%
Fish, Seafood	6.5%
Poultry Meat	6.2%
Pigmeat	5.6%
Vegetables, Other	4.7%
Bovine Meat	4.4%
Maize and products	4.4%
Eggs	3.4%

Source: FAO (2014).

Table 3.2: Foods ranked by contribution to world calories

FAO Food Balance Item	World Calorie Percentage
Rice (Milled Equivalent)	19%
Wheat and products	18.3%
Sugar (Raw Equivalent)	6.8%
Dairy	5.9%
Maize and products	5.1%
Pigmeat	4.2%
Soyabeans	3.4%
Vegetables, Other	2.5%
Potatoes and products	2.2%
Poultry Meat	2%

Source: FAO (2014).

Table 3.3: Land uses ranked by contribution to world food supply

FAO Food Balance Item	World Calorie Percentage including Contribution via Animals	
	Low	High
Rice (Milled Equivalent)	19.2%	19.1%
Wheat and products	18.9%	18.5%
Pasture and Crop Residues	9.7%	14.6%
Sugar (Raw Equivalent)	6.8%	6.8%
Maize and products	7.2%	6%
Soyabeans	3.5%	3.5%
Vegetables, Other	2.5%	2.5%
Potatoes and products	2.3%	2.3%
Palm Oil	1.8%	1.8%
Cassava and products	1.4%	1.4%

[1] The contribution of pasture and crop residues is not directly measured in the FAO food balance data. I fully describe the calculation of this item in Section B.1. The low estimate assumes that no pasture or crop residues are used to feed pigs and chickens and no grain is used to feed bovine animals, goats, and sheep. The high estimate assumes that all feeds are used with equal proportions for all animals.

[2] Sources: FAO (2014).

Table 3.3 shows estimates of the contribution of different land uses to the global food supply. I use animal contributions to the food supply, assumed feed conversion ratios, and data on the amount of grain feed utilized to calculate the contribution of pasture and crop residues as a residual. I fully describe the calculations in Section B.1. The "Low" estimate, meaning a low estimate for the contribution of pasture and crop residues to the food supply, assumes that no pasture or crop residues are used to feed pigs and poultry, both high feed conversion animals, and no grain is used to feed bovine animals, goats, and sheep. The high estimate assumes that all feeds are used with equal proportions for all animals. All food items that are also used as animal feed include both the contribution via direct consumption and via animals, not accounting for crop residues.

I find that, consistent with the prior two tables, that rice and wheat are the dominant contributors to the food supply, each with almost 20% of global caloric production. Pasture and crop residues is then the next highest category, providing 9.7% using the low estimate and 14.6% using the high estimate, with pasture contributing approximately 70% of the category (Wirsenius 2003). The contribution of pasture is then of a similar magnitude to that for maize, or possibly larger.

3.3 Data

Dairy production data

I use New Zealand dairy production data at the territorial local authority (TLA)[1] and dairy season[2] level from the New Zealand Dairy Statistics series published by Livestock Improvement Corporation (LIC) and DairyNZ, both industry-owned bodies. In this chapter, I will refer to TLAs as districts. The production statistics are compiled from raw data collected from all major New Zealand dairy companies and can be considered a near census of production.[3] The variables I use from these publications are the number of cows milked at least one day during the season, the number of farms, production of milk per farm, production of protein per farm, and production of milkfat per farm. Data are available annually from 1999 to 2015, and the dairy production data define the period of study for this chapter.

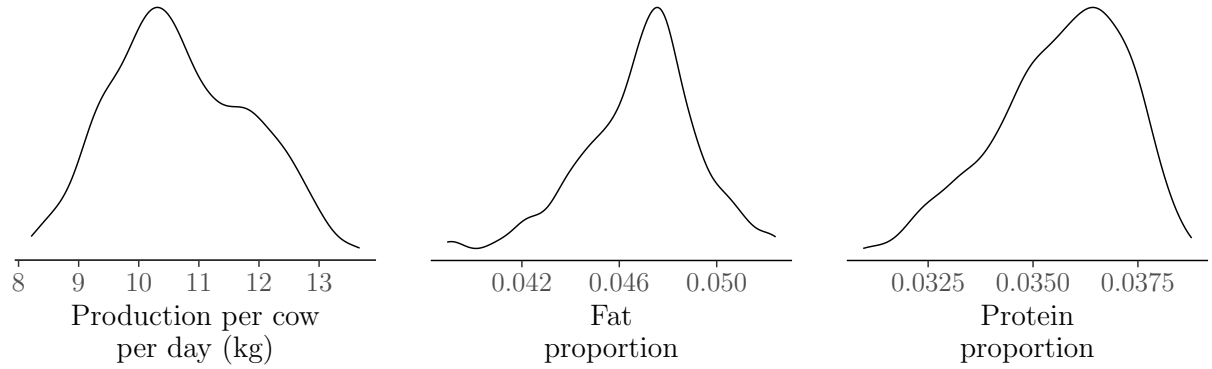
Figure 3.1 shows the distributions of each of the main outcome variables of interest in this chapter, the yield of milk per cow per day, the proportion of milkfat, and the proportion of protein.

[1] Similar to US counties or cities.

[2] In New Zealand, this is June 1 to May 31.

[3] A very small amount of boutique local supply milk is not counted in these data.

Figure 3.1: Density plots of the distributions of dairy outcome variables. Annual LIC data by district from New Zealand from 1999-2015.



Weather data

New Zealand's primary atmospheric research unit, the National Institute of Atmospheric Research (NIWA) provides a gridded weather data product, the Virtual Climate Station Network (VCSN), which provides a rich array of weather variables on a daily scale and on a regular grid of approximately 5km × 5km. The variables I use from the VCSN are daily rainfall, maximum and minimum air temperature, and soil moisture. Following Schlenker and Roberts (2009), I interpolate minimum and maximum air temperature using the single sine method, and compute linear spline transformations of all variables.

I obtain the 2015 proportion of each VCSN grid cell that is covered by dairy land from the Ministry of Primary Industries. I use these proportions to weight the weather grid cells when aggregating spatially. I aggregate the dairy-land-weighted weather grid cells using a dataset of district polygons using area-overlap weights.

All transformations are first done at the grid cell-day level, then aggregated spatially and temporally. I include distributions of all included variables in the Results section.

The soil moisture variable used in this chapter is modeled from a time series of historical temperature and precipitation data as described in Porteous, Basher, and Salinger (1994). These authors report that the model yields results with quite high accuracy, though I do not know of more recent validation exercises.

3.4 Conceptual framework

In this chapter, the big-picture parameter of interest is the change in producer surplus in a local production center under climate change. Local production regressions that use historical data are able to recover information about the production function as it relates to weather variables; however, these regressions are unable to be used to estimate changes in

surplus under climate change due to induced changes in either input prices, output prices, or management actions which are not available on the same time-scale of the shocks used to estimate the regression parameters (Hsiang 2016). While the data used in this chapter are not rich enough to give indications of potential changes in input prices, output prices, or long-run management changes, I would like to be explicit that these are missing pieces to be filled in by future research, as is the potential impact on consumer surplus, where several downstream industries (processing, logistics, and retail), complicate a full accounting. In this section, I formalize these ideas using a standard production function framework that considers a final consumption good produced in a competitive market, where environmental conditions enter the production function directly, as well as affecting the prices of both inputs and output.

Production of the good in the local region, y, depends on inputs that are adjustable in the short-run, $\boldsymbol{X}_s$, inputs that are fixed in the short-run, $\boldsymbol{X}_l$, and local environmental conditions $\boldsymbol{E}_\ell$. Examples of inputs that are adjustable in the short-run are fertilizer, water use, and use of off-farm feed, while examples of inputs that are fixed in the short-run are milking infrastructure, cooling infrastructure, and on-farm feed species.

$$y = f(\boldsymbol{X}_s, \boldsymbol{X}_l, \boldsymbol{E}_\ell) \tag{3.1}$$

Given the assumption that there are global markets for all commodities, prices of both inputs and outputs will depend on environmental conditions in all locations the goods are produced in, which I denote as $\boldsymbol{E}$, with $\dim(\boldsymbol{E}_\ell) < \dim(\boldsymbol{E})$. Profit, π, for the competitive firm, is then:

$$\pi = p_y(\boldsymbol{E}) f(\boldsymbol{X}_s, \boldsymbol{X}_l, \boldsymbol{E}_\ell) - \boldsymbol{p}_{\boldsymbol{X}_s}(\boldsymbol{E}) \cdot \boldsymbol{X}_s - \boldsymbol{p}_{\boldsymbol{X}_l}(\boldsymbol{E}) \cdot \boldsymbol{X}_l \tag{3.2}$$

The firm's optimization implies that optimal output, y^*, depends on both local and outside environmental conditions:

$$y^* = f(\boldsymbol{X}_s^*(\boldsymbol{E}_\ell, \boldsymbol{E}_{-\ell}), \boldsymbol{X}_l^*(\boldsymbol{E}_\ell, \boldsymbol{E}_{-\ell}), \boldsymbol{E}_\ell) \tag{3.3}$$

where $(\boldsymbol{E}_\ell, \boldsymbol{E}_{-\ell}) \equiv \boldsymbol{E}$. In this framework, I represent global climate change as a change in the full vector $\boldsymbol{E}$, denoted by $\Delta\boldsymbol{E}$. To simplify the exposition, I make the assumption that prices do not depend on supply in the local region: $p_y(\boldsymbol{E}) = p_y(\boldsymbol{E}_{-\ell})$, $\boldsymbol{p}_{\boldsymbol{X}}(\boldsymbol{E}) = \boldsymbol{p}_{\boldsymbol{X}}(\boldsymbol{E}_{-\ell})$, where $\boldsymbol{p}_{\boldsymbol{X}} \equiv (\boldsymbol{p}_{\boldsymbol{X}_s}, \boldsymbol{p}_{\boldsymbol{X}_\ell})$.

Omitting cross terms, the change in production is then:

$$\begin{aligned}\Delta y^* \approx \quad & \underbrace{\frac{\Delta f}{\Delta \boldsymbol{E}_\ell} \cdot \Delta \boldsymbol{E}_\ell}_{\text{Direct production effect}} + \underbrace{\frac{\Delta f}{\Delta \boldsymbol{X}_s} \cdot \frac{\Delta \boldsymbol{X}_s^*}{\Delta \boldsymbol{E}_\ell} \cdot \Delta \boldsymbol{E}_\ell}_{\text{Short-run adaptation effect}} + \underbrace{\frac{\Delta f}{\Delta \boldsymbol{X}_l} \cdot \frac{\Delta \boldsymbol{X}_l^*}{\Delta \boldsymbol{E}_\ell} \cdot \Delta \boldsymbol{E}_\ell}_{\text{Long-run adaptation effect}} + \\ & \underbrace{\frac{\Delta f}{\Delta \boldsymbol{X}} \cdot \frac{\Delta \boldsymbol{X}^*}{\Delta p_y} \cdot \frac{\Delta p_y}{\Delta \boldsymbol{E}_{-\ell}} \cdot \Delta \boldsymbol{E}_{-\ell}}_{\text{Output price effect}} + \underbrace{\frac{\Delta f}{\Delta \boldsymbol{X}} \cdot \frac{\Delta \boldsymbol{X}^*}{\Delta \boldsymbol{p}_{\boldsymbol{X}}} \cdot \frac{\Delta \boldsymbol{p}_{\boldsymbol{X}}}{\Delta \boldsymbol{E}_{-\ell}} \cdot \Delta \boldsymbol{E}_{-\ell}}_{\text{Input price effect}}\end{aligned} \tag{3.4}$$

where $\boldsymbol{X} \equiv (\boldsymbol{X}_s, \boldsymbol{X}_l)$. Equation (3.4) highlights that long run environmental shocks can affect output through several channels. The first term represents the direct impact climate has on production; in this context through changes in on-farm feed growth and heat stress. The second term represents the impact of climate change on production through adjustments in short-term inputs. In the regression framework, the sum of these first two terms is able to estimated with a standard fixed effects approach, which this chapter implements. With a long dataset, researchers can estimate the sum of the first three terms using a long-differences approach (Burke and Emerick 2016), which involves a trade-off in terms of statistical power (Hsiang and Burke 2013). However, the estimation of the final two terms depends on global market responses to climate change, so is not able to be informed by data on production in a single area.

In the New Zealand dairy context, because such a large proportion of feed comes from on-farm, I would like readers to interpret $\Delta f / \Delta \boldsymbol{E}_\ell$, the direct effect of the environment, as being the combined effect of the environment directly on cows as well as via shocks to local feed. In the extreme, the supplemental feed portion of $\boldsymbol{X}_s$ is then at a corner solution where feed consumed is equal to zero, which is often the case in New Zealand. In this context, I am able to provide information on the response of pasture to environmental shocks, via the effect on dairy production.

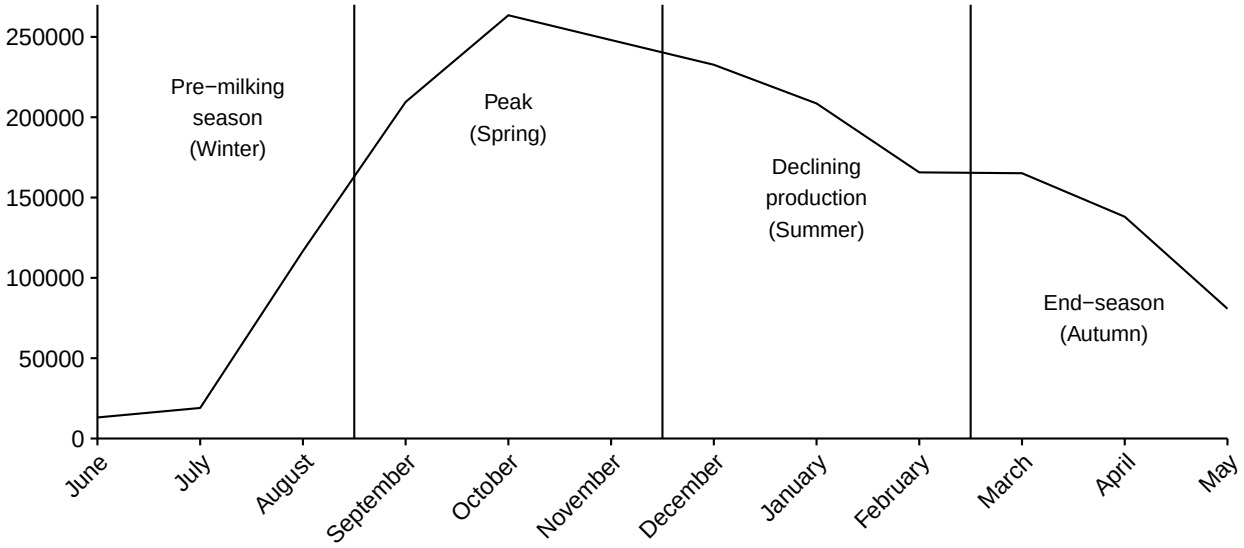
When focusing on the level of production, y, we expect that the direct effects will dominate the climate change induced changes in production. However, as aforementioned, the big picture quantity of interest for this line of research is the change in producer surplus, which also includes first-order changes in both output and input prices. Omitting both cross terms and second-order terms, the change in producer profit is:

$$\Delta \pi^* \approxeq \underbrace{\frac{\Delta f}{\Delta \boldsymbol{E}_\ell} \cdot p_y \cdot \Delta \boldsymbol{E}_\ell}_{\text{Production effect}} + \underbrace{\frac{\Delta p_y}{\Delta \boldsymbol{E}_\ell} \cdot y^* \cdot \Delta \boldsymbol{E}_\ell}_{\text{Output price effect}} + \underbrace{\frac{\Delta \boldsymbol{p}_{\boldsymbol{X}}}{\Delta \boldsymbol{E}_{-\ell}} \cdot \boldsymbol{X}^* \cdot \Delta \boldsymbol{E}_{-\ell}}_{\text{Input price effect}} \tag{3.5}$$

Yield studies like this chapter contribute to the estimation of the first term in Equation (3.5), but a global analysis would be required to estimate the second and third terms, which is a task for future research.

In summary, the fixed effects models that I show in this chapter are able to estimate the sum of the direct production effect of weather via thermal stresses on cows, the direct production effect of weather via impacts on locally grown feed, and the indirect production effect of weather via changes in inputs that are adjustable in the short-term. Output and input prices have a first-order effect on producer surplus but are not included in the estimation in this chapter.

Figure 3.2: New Zealand production by month in 2015-2016



3.5 Methods

This chapter uses established econometric methods to estimate the relationship between weather and my economic variable of interest, the average milk yield per cow, within a district. For a review of the econometrics of weather and climate, see Hsiang (2016). Following Schlenker and Roberts (2009), for my primary specifications I use linear splines, with knot locations estimated using nonlinear least-squares (NLLS). I can express each of the estimating equations in the following form:

$$y_{it} = f(\boldsymbol{T}_{it}) + g(\boldsymbol{M}_{it}) + \delta_{i0} + \delta_{i1}t + \varepsilon_{it} \tag{3.6}$$

where y_{it} is dairy production outcome of interest for district i and year t. In this chapter, y_{it} is one of the average milk yield per cow, the average protein content of milk, or the average fat content of milk, each within a district. f is a function of the full vector of daily temperature within a district-year ($\boldsymbol{T}_{it}$), g as a function of the full vector of daily soil moisture within a district-year ($\boldsymbol{M}_{it}$), and $\delta_{i0} + \delta_{i1}t$ is a unit specific affine trend.

In all specifications in this chapter, f is a linear spline function in temperature interpolated using the single sine method and g is is a linear spline function in daily soil moisture, each with a single knot. I estimate each specification using NLLS, restricting all knot locations to be between the 10th and 90th percentiles of daily support of the regressors.

Figure 3.2 shows country-level monthly production in the 2015-2016 dairy season. The plot clearly shows the seasonality in dairy production in New Zealand. Importantly, the seasonality motivates me to relax the assumption in the standard model that a weather shock in different stages of the dairy season has the same impact. One might expect that a

weather shock in spring would have a larger effect on annual production than an equivalent shock in autumn, both because there is more production to impact, and a longer amount of time remaining in the aggregation period for lagged effects to realize.

In specifications that restrict the response to weather to be the same throughout the year, I aggregate the weather variables across the New Zealand dairy season (June-May). In specifications that allow for flexibility in weather responses by time-of-year, I aggregate the weather variables by weather seasons, June-August, September-November, December-February, and March-May, which also corresponds to general stages of the New Zealand dairy season.[4]

3.6 Results

Figure 3.3 plots the temperature response function in a regression with no control variables and aggregating the temperature variables across the entire dairy season. The red dashed lines indicate positive and negative average daily production; I include these to give readers a sense of the scale of the impacts. The NLLS estimator finds an optimal turning point of 19°C, with a statistically insignificant effect of increasing temperatures below the turning point and a large and statistically significant negative effect of increasing temperatures above the turning point. To give a sense of the magnitude of the estimated relationship, moving 24 hours of temperature from 19°C to 22°C reduces annual production by the equivalent of that for an average day of the year, a large effect.

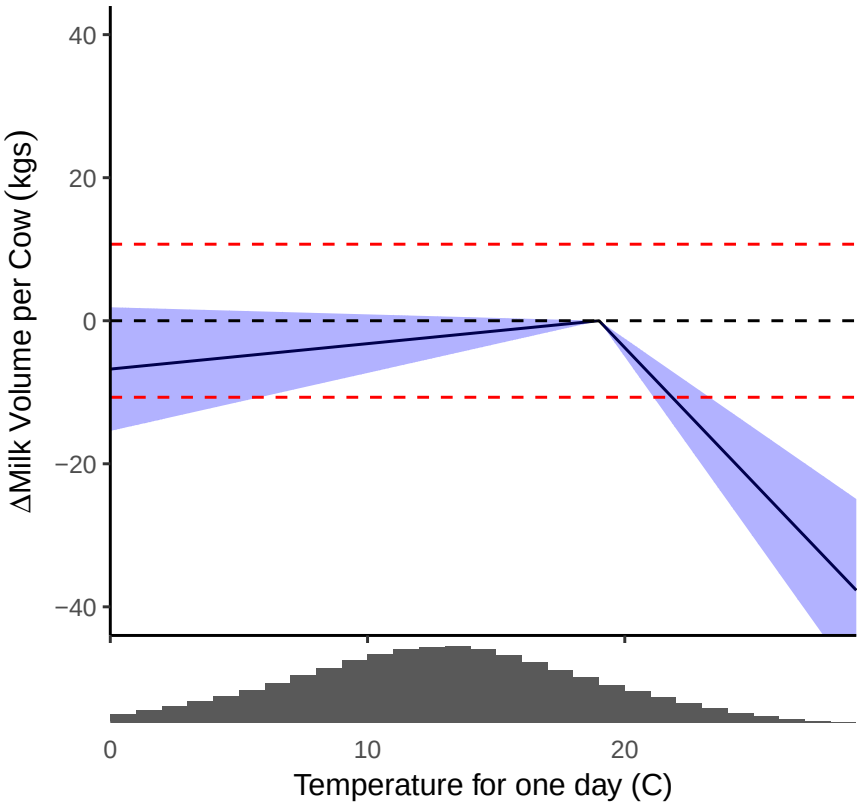
Figure 3.4 plots the temperature response function introducing soil moisture as a control variable. In this response function, one would have to move 24 hours of temperature from 19°C to 24°C to reduce annual production by the equivalent of that for an average day of the year, showing that some of the temperature effect works through its impact on soil moisture. This is a substantial reduction in the direct effect of temperature but it remains large.[5] In unreported results, I find that including precipitation does not provide extra information over the soil moisture variable and, thus, I omit it.

Figure 3.5 plots the soil moisture response function for the specification where weather is aggregated across the entire dairy season. The NIWA data represents soil moisture as the

[4]In financial econometrics, the flexible specification is known as the "step function MIDAS" model (Ghysels, Sinko, and Valkanov 2007). This literature examines the econometrics of regressions between variables of differing frequencies. Here, for example, I am regressing *annual* milk yields on *daily* weather. The MIDAS literature is mostly focused on estimators that allow flexible specification of the structure of how response functions change throughout the aggregation period. I use the "step function MIDAS" model, where the aggregation is flat within several subperiods of the low frequency variaable, as it allows me to use OLS estimation, after the estimation of linear spline knot locations using NLLS.

[5]Bell (2017, Chapter 2), in the context of maize in the United States, gives suggestive evidence that the estimated effect of extreme temperature could be partially due to mis-measurement of precipitation. It is possible that the estimation in this chapter suffers from the same problem described in Bell. A future iteration of this project will implement the solution suggested in that chapter.

Figure 3.3: Production response function to temperature with no controls

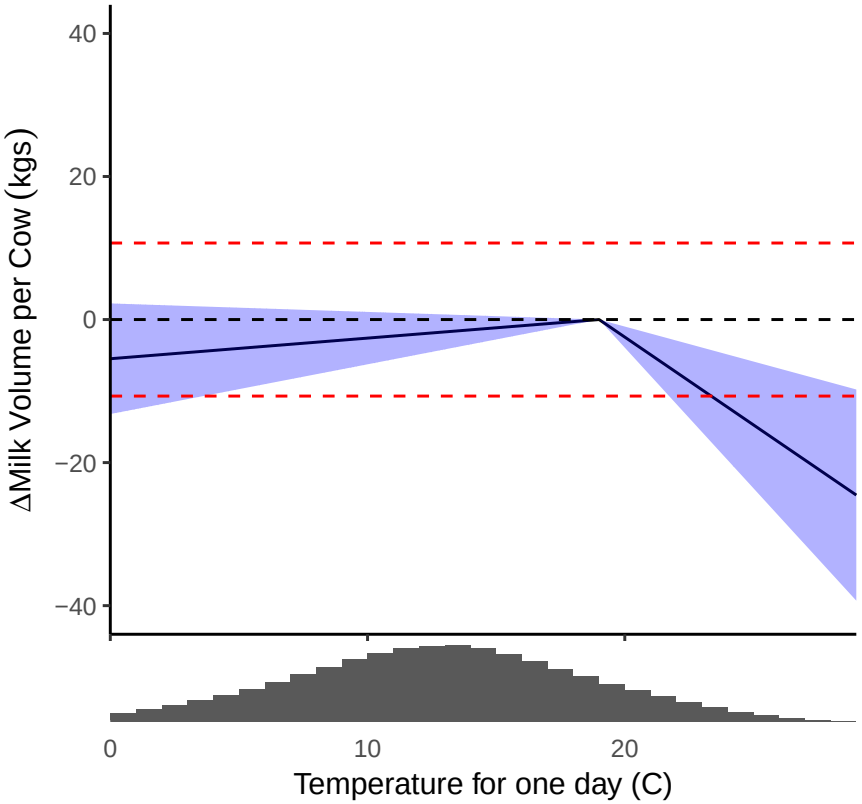


This figure plots f in Equation (3.6) with $g = 0$ for a single day of temperature, estimated using NLLS. The turning point is restricted to lie between the 10th and 90th percentiles of the full distribution of temperature. The plot is vertically centered so that the change in yield per cow takes a value of zero when temperature is 19° C. The dashed red lines indicate positive and negative average daily production. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the plot is a histogram of the full time series of temperature using single sine interpolation.

negative of the "soil moisture deficit", which is the quantity of rainfall required to bring the soil up to capacity. In these units, very dry soils have large negative numbers and very wet soils have positive numbers. The figure finds that extreme wet conditions have large negative impacts on milk yields, and that dry conditions have moderate negative impacts. The main mechanism by which very wet soils impact milk production in the pasture-based context is via soil compaction. Compaction occurs when stock intensively trample wet soil, causing poor subsequent water absorption and lower pasture growth. I find that the overall scale of yield impacts of changes in soil moisture is large but smaller than that of temperature.

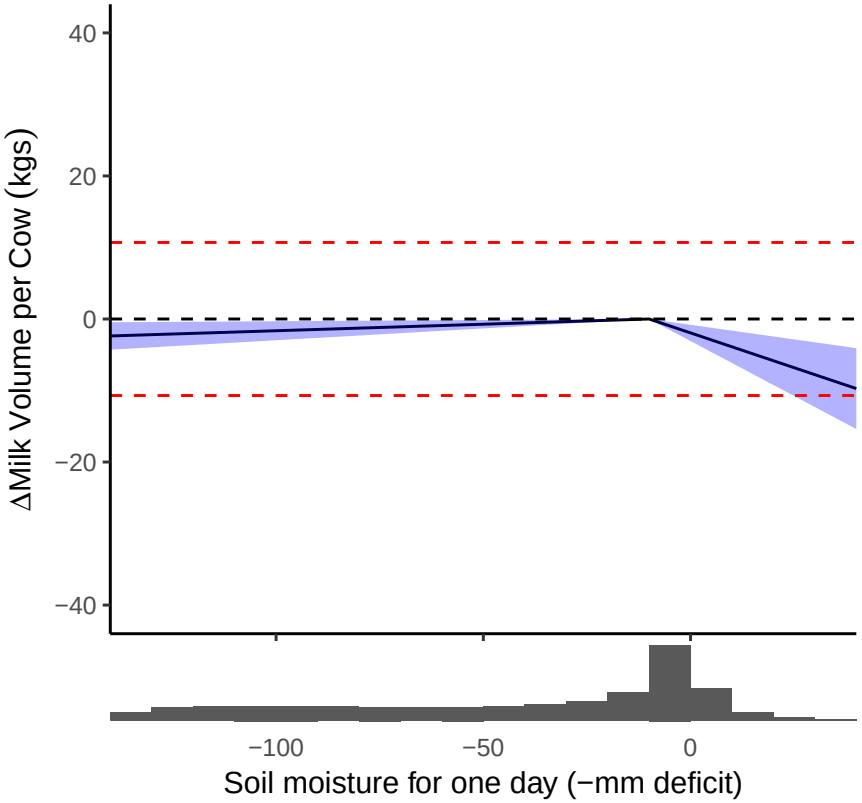
Figure 3.6 shows the temperature response functions estimated separately for the different weather seasons, including soil moisture controls. Several qualitative differences emerge when moving to this more flexible specification. Firstly, I find large and positive impacts

Figure 3.4: Production response function to temperature with soil moisture controls



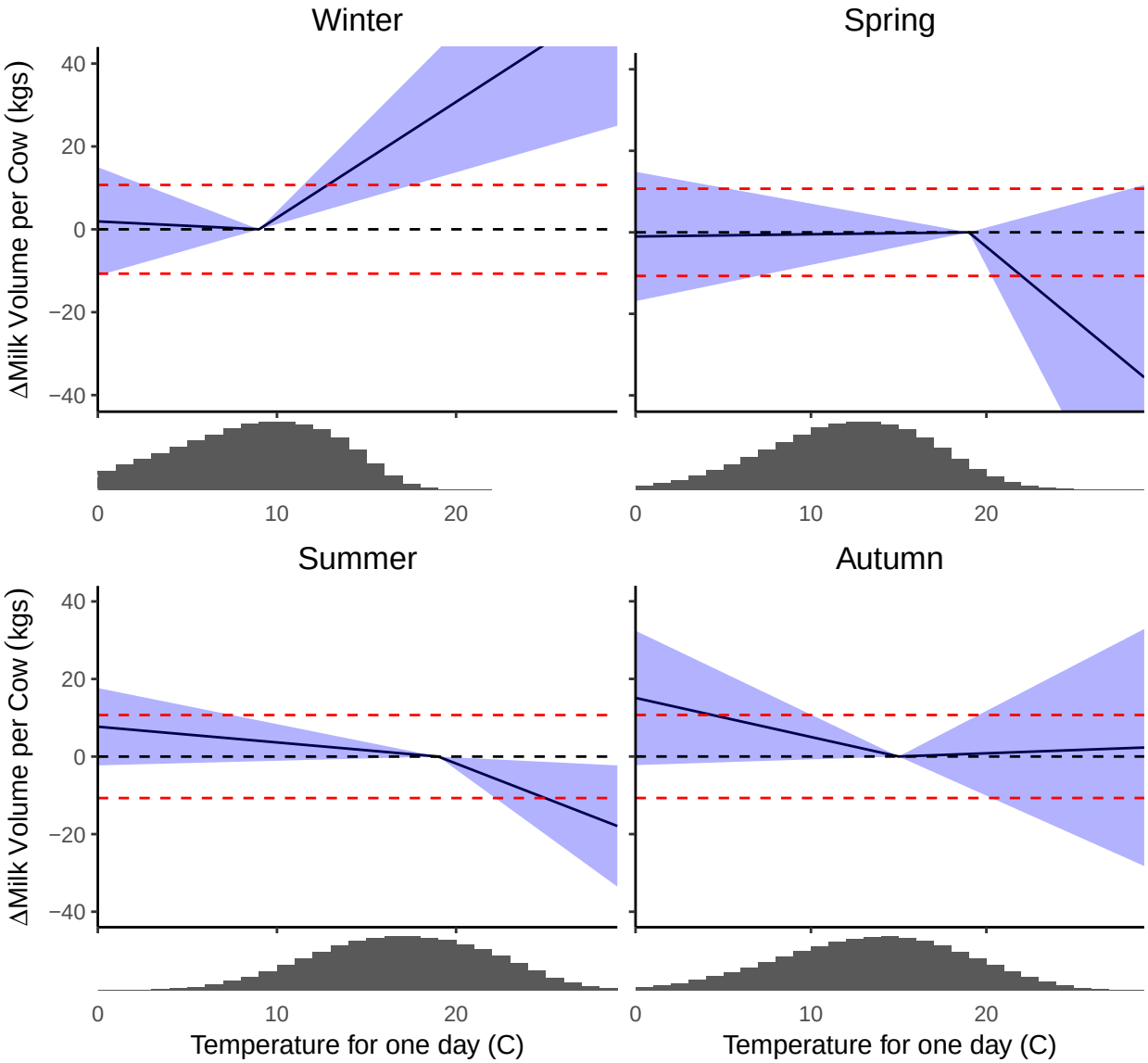
This figure plots f in Equation (3.6) for a single day of temperature, estimated using NLLS. The turning point is restricted to lie between the 10th and 90th percentiles of the full distribution of temperature. The plot is vertically centered so that the change in yield per cow takes a value of zero when temperature is 19° C. The dashed red lines indicate positive and negative average daily production. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the plot is a histogram of the full time series of temperature using single sine interpolation.

Figure 3.5: Production response function to soil moisture



This figure plots g in Equation (3.6) *for a single day of soil moisture, estimated using NLLS. The turning point is restricted to lie between the 10th and 90th percentiles of the daily distribution of soil moisture. In the NIWA data, the units of soil moisture are the negative of the quantity of water in mm required to bring the soil up to capacity; this plot uses these units. Positive values indicate the quantity of water running off. The plot is vertically centered so that the change in yield per cow takes a value of zero when the negative of soil moisture deficit is -10mm. The dashed red lines indicate positive and negative average daily production. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the plot is a histogram of daily soil moisture.*

Figure 3.6: Production response function to temperature estimated by weather season



This figure plots f in Equation (3.6), where the linear spline functions are estimated by weather season, for a single day of temperature, estimated using NLLS. The subplots are vertically centered so that the change in yield per cow takes a value of zero when temperature is at the respective knot locations. The dashed red lines indicate positive and negative average daily production, averaged over the year. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the subplots are histograms of the full time series of temperature, within a weather season, using single sine interpolation.

of moderate temperatures during the pre-dairy-season winter. This is likely due to pasture responses that both result in improved cow condition before calving and larger pasture stocks for spring grazing. Secondly, I find negative and marginally insignificant impacts of moderate temperature in summer and autumn. Thirdly, I find no impact of extreme temperature on milk yields in autumn.

Figure 3.7 plots the soil moisture response functions by weather season. As in Figure 3.5, I find small impacts of drying soils and I find that the large negative impact of very wet soils is concentrated in spring. In spring, stock are on pasture much more than in winter, so we would expect to see the negative effects of compaction much more in spring.

Figure 3.8 plots the response functions of fat proportion to temperature, estimated by season. I choose the knot locations to be the same as in the milk volume regression. Though economically smaller than the impacts on milk volume,[6] I find negative impacts of increasing temperatures across the spectrum in summer, spring, and autumn, with these impacts statistically significant for moderate summer temperatures. I find large negative impacts of increasing cool temperatures in winter and large positive impacts of increasing moderate temperatures in winter. Figure 3.9 plots the same for protein proportion and finds the same qualitative and similar quantitative results as Figure 3.8.

Projections under climate change

To more closely examine the economic significance of the results of the previous subsection, I use the HadGEM-ES climate model to project forward changes in milk production under climate change. Following Burke, Dykema, et al. (2014), I simulate future weather by adding changes from the climate model to historical weather levels. As in Houser et al. (2015), I randomly choose historical weather years to map to future simulated years.

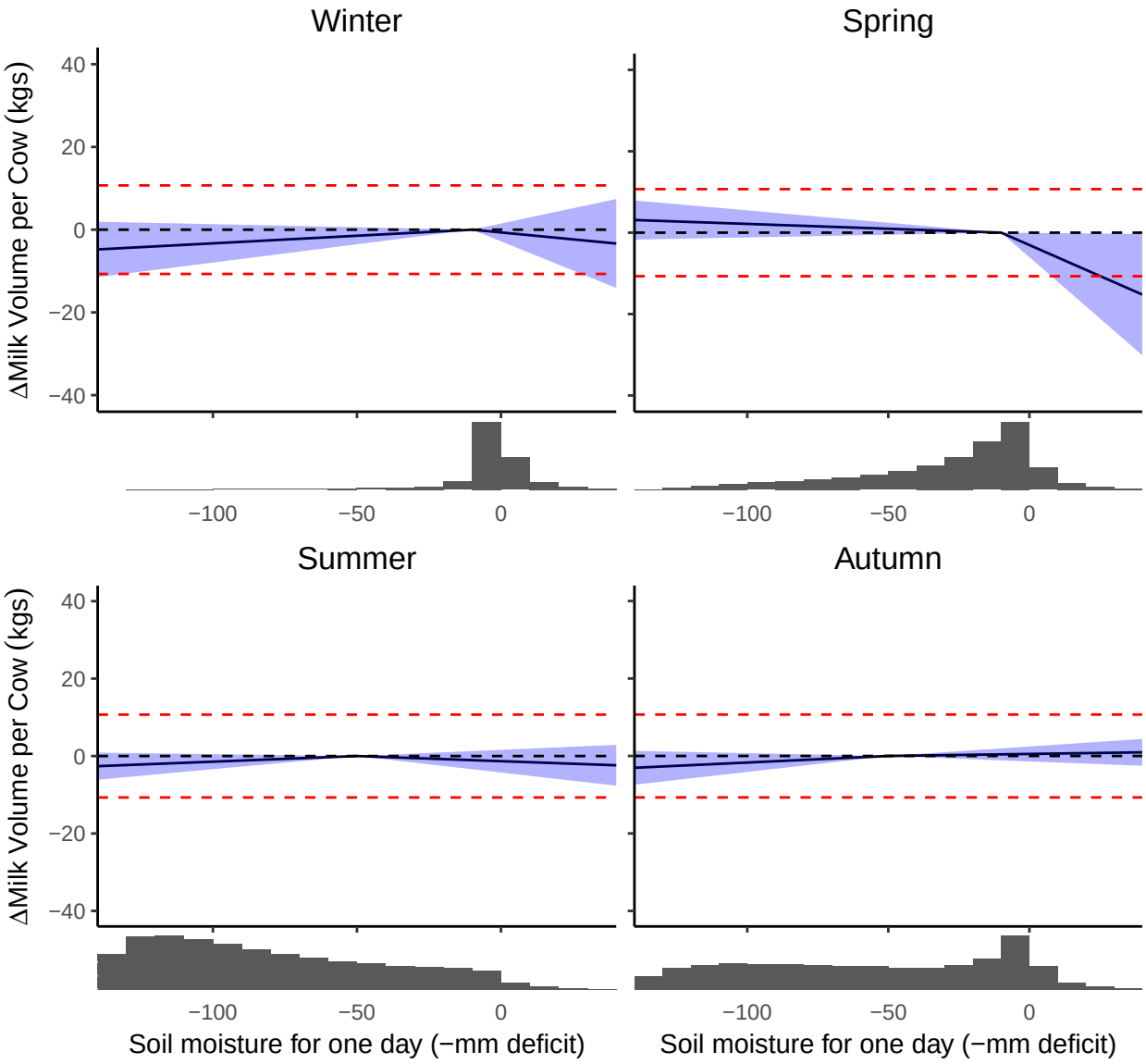
Unlike past work, I compute projections for all future years to 2100. To isolate only decadal variation from the climate model, I compute LOWESS smoothed trends of each variable by month-of-year. The full details of the projection computation is in Section B.2.

Figure 3.10 plots the projection results as the proportion of lost annual revenue under climate change. The main stark feature from this figure is that the model that restricts response functions to be the same throughout the year results in much more pessimistic forecasts of the response to climate change. While the two projections trend downwards from around 2035, the flexible model both reduces ultimate production by less and includes a larger upward trend at the beginning of the simulation period, which is statistically significant.

To better show the relative economic importance of the initial upward trend to around 2035 versus the subsequent downward trend, Figure 3.11 plots the same data as the previous figure adjusting all quantities with a 3% discount rate. The proportions plotted in this figure are then discounted revenues as a proportion of current revenue. This figure make clear that

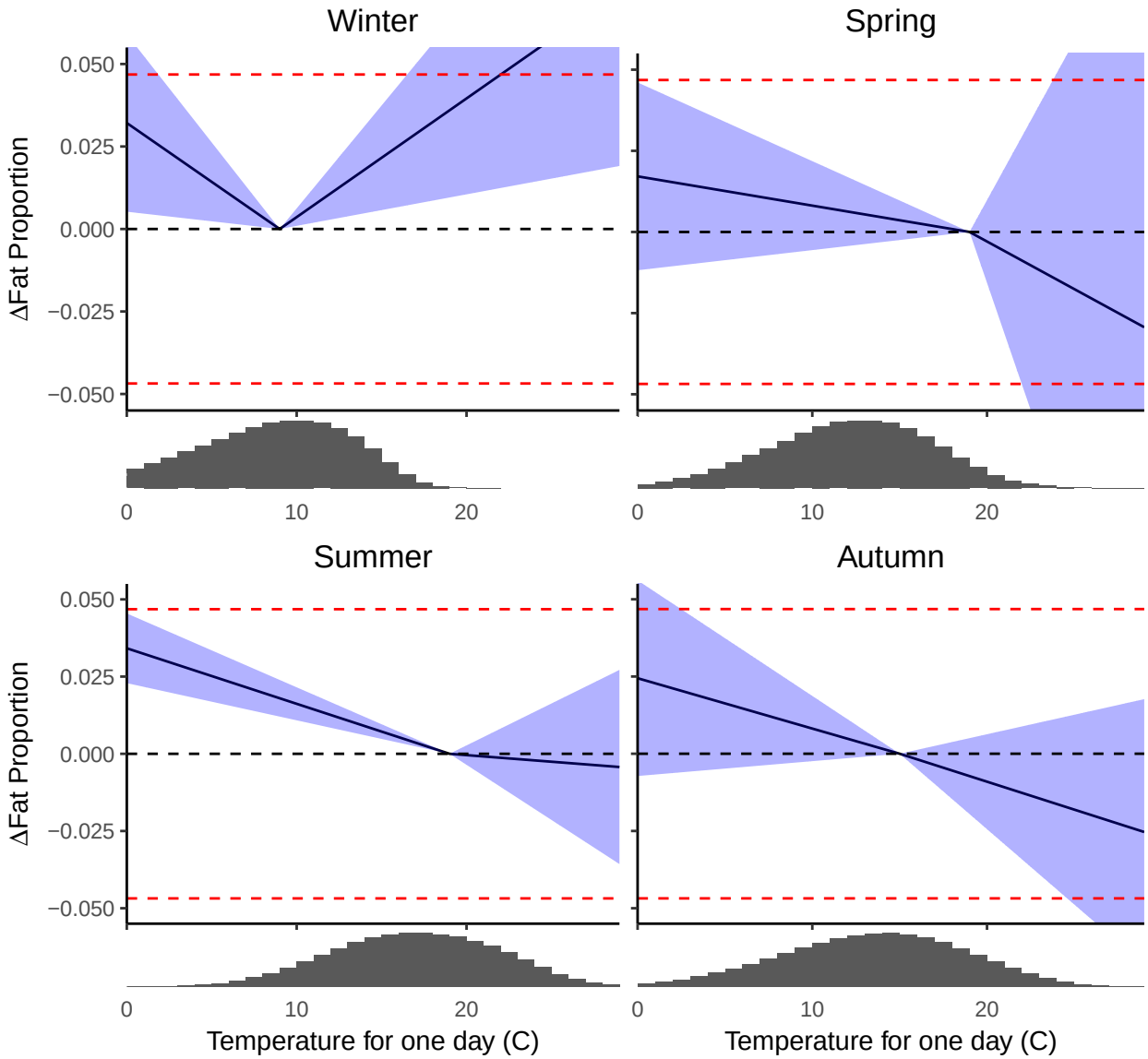
[6]Note that the red dashed lines indicating average daily fat proportion are placed closer to the limits of the plots.

Figure 3.7: Production response to soil moisture estimated by weather season



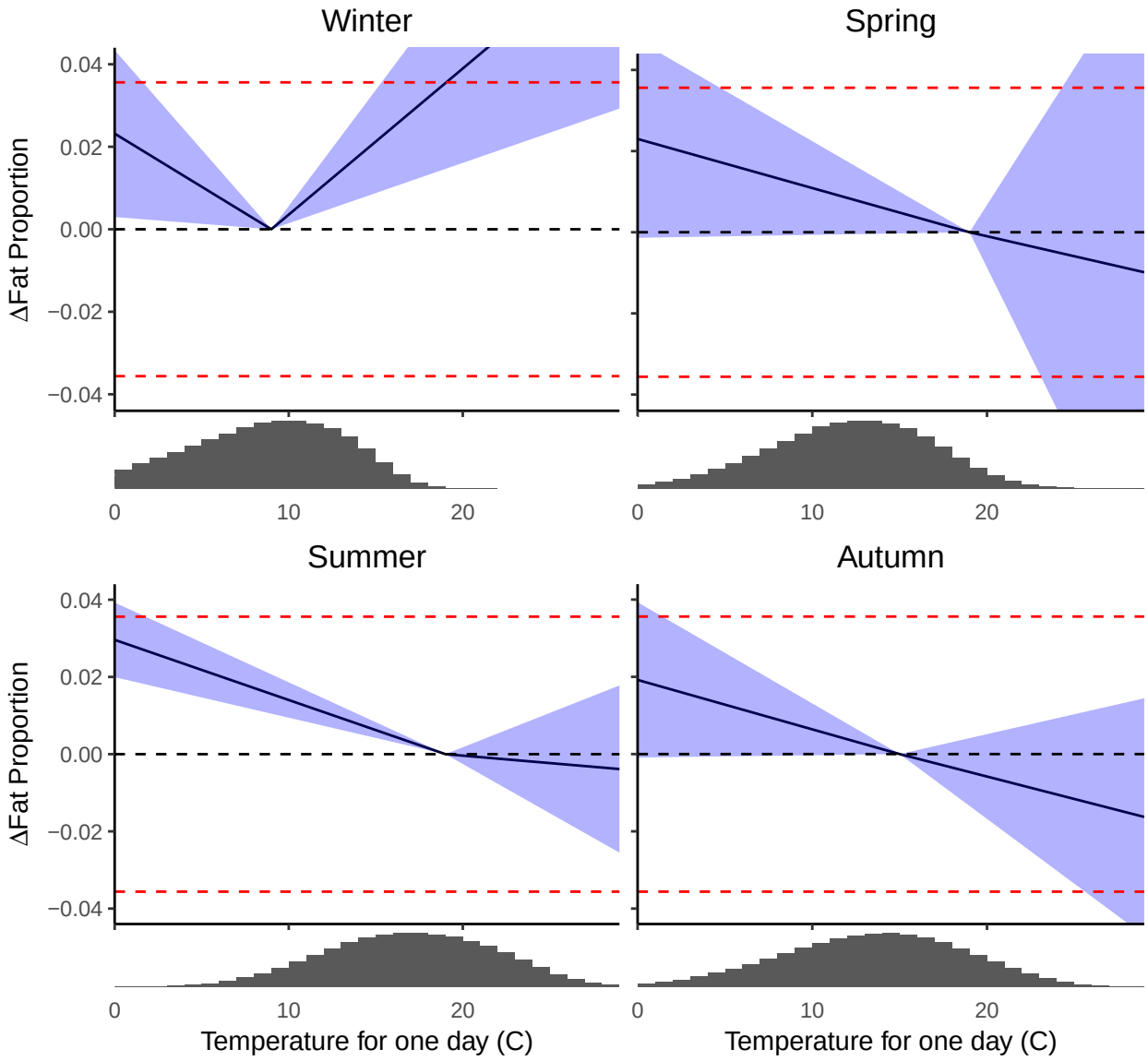
This figure plots g in Equation (3.6)*, where the linear spline functions are estimated by weather season, for a single day of soil moisture. In the NIWA data, the units of soil moisture are the negative of the quantity of water in mm required to bring the soil up to capacity; this plot uses these units. Positive values indicate the quantity of water running off. The subplots are vertically centered so that the change in yield per cow takes a value of zero when the negative of soil moisture deficit is at the respective knot locations. The dashed red lines indicate positive and negative average daily production. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the subplots are histograms of daily soil moisture, within a weather season.*

Figure 3.8: Fat response function to temperature estimated by weather season



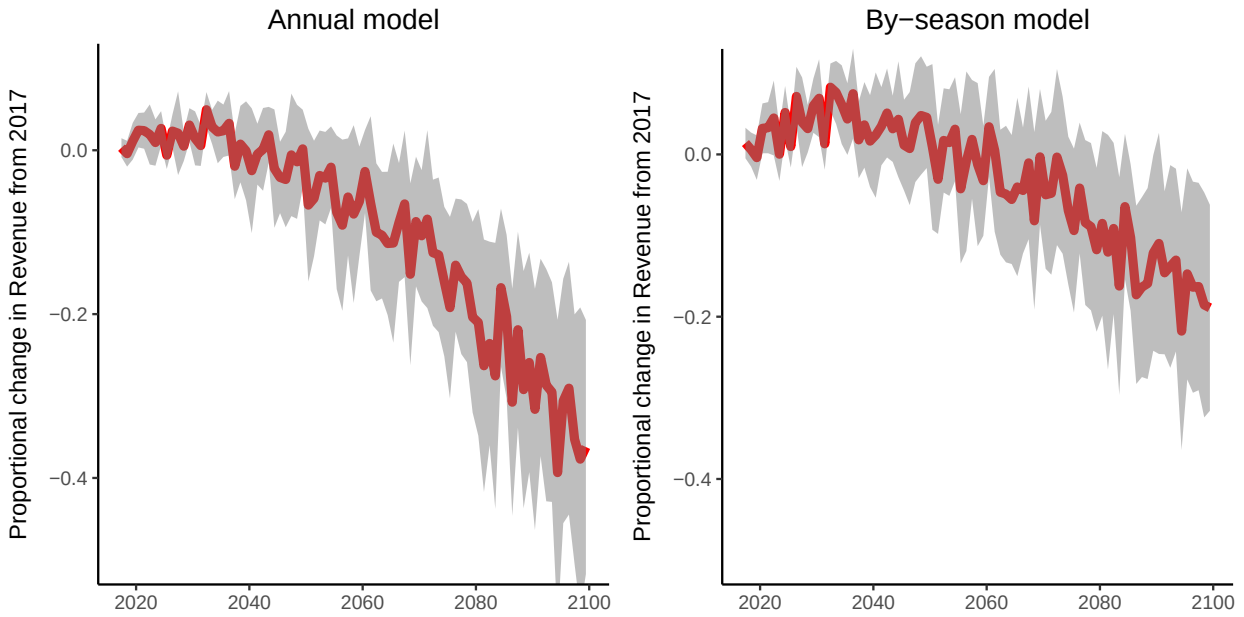
This figure plots f in Equation (3.6), where the linear spline functions are estimated by weather season, for a single day of temperature, using fat proportion as the outcome variable. I use the same knot locations estimated in the milk yield regressions. The subplots are vertically centered so that the change in yield per cow takes a value of zero when temperature is at the respective knot locations. The dashed red lines indicate positive and negative average fat proportion, averaged over the year. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the subplots are histograms of the full time series of temperature, within a weather season, using single sine interpolation.

Figure 3.9: Protein response function to temperature estimated by weather season



This figure plots f in Equation (3.6), where the linear spline functions are estimated by weather season, for a single day of temperature, using protein proportion as the outcome variable. I use the same knot locations estimated in the milk yield regressions. The subplots are vertically centered so that the change in yield per cow takes a value of zero when temperature is at the respective knot locations. The dashed red lines indicate positive and negative average fat proportion, averaged over the year. The plot shows 95% confidence bands calculated assuming error clustering by district and year; confidence bands do not take account of the uncertainty in the knot location. Below the subplots are histograms of the full time series of temperature, within a weather season, using single sine interpolation.

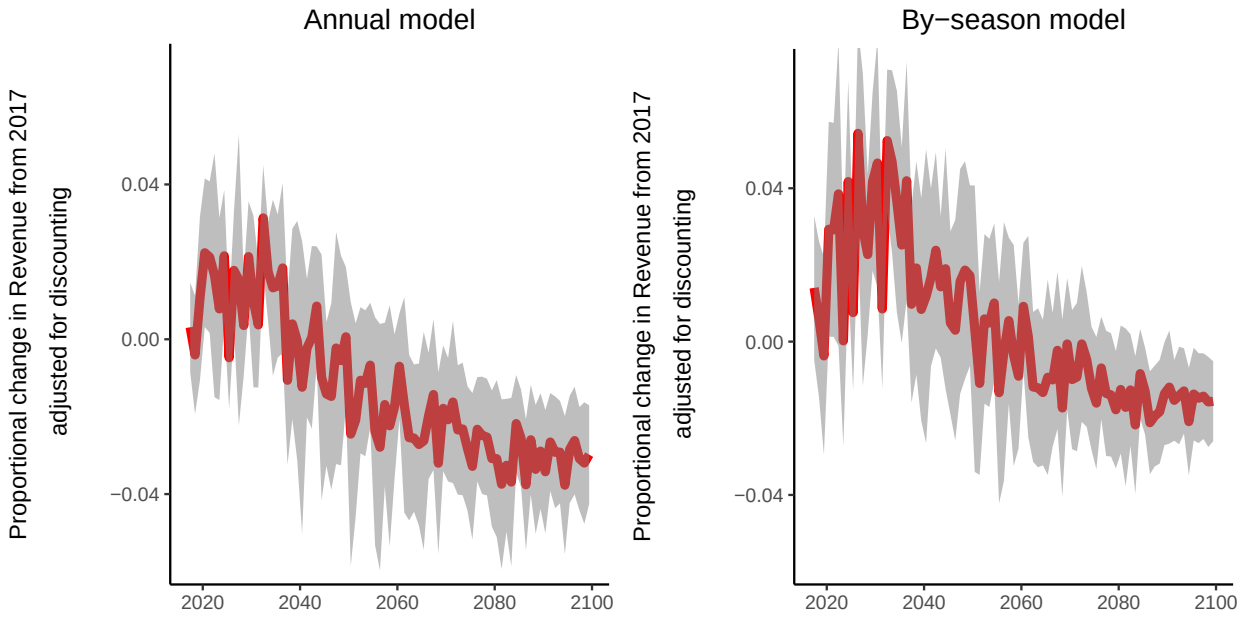
Figure 3.10: Projected change in revenue



This figure plots the projected proportional change in revenue under climate change as simulated in the HadGEM2-ES model for the annual and the by-season models. The current value of annual New Zealand dairy production at the farm gate is approximately US$8 billion. The projection simulation assumes constant output prices and constant milk quality. The simulation projects both temperature and soil moisture forward. At a 3% discount rate, the point estimate of the present value of the change in production is -US$8.4 billion for the annual model, and +US$2.1 billion for the by-season model.

the negative impacts in the later period dominate the projection results in terms of economic importance in the restricted annual model, whereas the early period gains balance with the later period losses in the flexible by-season model. When aggregating these values, I find that the point estimate of the present value of the change in production is -US$8.4 billion for the annual model, and +US$2.1 billion for the by-season model. The current value of annual New Zealand dairy production at the farm gate is approximately US$8 billion.

Figure 3.11: Projected change in revenue with discounting



This figure plots the projected proportional change in revenue, adjusted with a 3% discount rate, under climate change as simulated in the HadGEM2-ES model for the annual and the by-season models. The current value of annual New Zealand dairy production at the farm gate is approximately US$8 billion. The projection simulation assumes constant output prices and constant milk quality. The simulation projects both temperature and soil moisture forward. The point estimate of the present value of the change in production is -US$8.4 billion for the annual model, and +US$2.1 billion for the by-season model.

3.7 Conclusion

This chapter has estimated the impact of weather variables on dairy production in New Zealand. It finds that restricting the weather response functions to be the same throughout the year results in more pessimistic projections of the consequences of future climate change in the New Zealand dairy context than allowing for responses to be flexible by time-of-year. More generally, it highlights that bias can result in applied research contexts with dependent variables of a lower frequency than independent variables and models that restrict responses to be the same throughout the aggregation period, as has been shown theoretically (Andreou, Ghysels, and Kourtellos 2010).

It also highlights that pasture-based livestock production is highly sensitive to the weather, with large and opposing effects of winter and summer temperatures. New Zealand dairy production exists in a temperate climate, with temperatures seldom moving outside the range of 0-30°C. If these results are indicative of the weather-pasture production relationship in cooler or warmer places, they imply that these areas will respectively experience large gains and large declines under climate change.

While New Zealand dairy producers are highly exposed to the global export industry, an important stylized fact about the industry more generally is that production tends to be close to consumption. While a global market exists for milk powder, cheese, butter, and whey, exports only account for around 10% of global milk production. Fluid milk, in particular, has very high transportation costs both due to the water carrier needing to be transported and spoilage. If my results are indicative of the weather-dairy production relationship in other contexts, this fact suggests that a large portion of the incidence of the costs and benefits of climate change will fall on consumers.

Chapter 4

The potential for renewable fuels under greenhouse gas pricing: The case of sugarcane

4.1 Introduction

Oil accounts for 33% of global greenhouse gas (GHG) emissions (International Energy Agency 2015). It is crucial to examine the scaling potential of alternative fuels as ambitious climate mitigation action is considered around the world, since the costs of possible policies, such as cap and trade programs and carbon taxes, must be evaluated. In the 1970s, Brazil established the first large-scale alternative transportation fuel sector under the Pró-Álcool program, demonstrating the commercial viability of the ethanol industry. The sector expanded aggressively in the 2000s; however, in recent years, a constellation of factors has contributed to slowing investment in new capacity. These include an unfavorable policy environment, particularly restrictions on the ownership of land that discourage the entry of foreign capital, energy policies that encouraged expansion of the oil sector, especially by investing in deepwater production (Moraes and Zilberman 2014), and decreasing energy prices. Today, ethanol is mandated to be blended into domestic fuel in Brazil at a proportion of 27%, providing an implicit subsidy to producers.

Three features of the Brazilian ethanol context make it particularly attractive to consider the expansion of low-carbon fuels on a large scale. First, the Brazilian climate allows for high yields of sugarcane (both in potential and reality), which is readily convertible into ethanol using a production process that emits relatively little GHG, when compared to gasoline or US corn ethanol (State of California 2009). Brazil also has a large amount of pasture land that is appropriate for growing this sugarcane (approximately 170 Mha, compared to around 10 Mha in use today), which emits a relatively small amount of carbon dioxide when converted to sugarcane, compared with forest or savanna systems (Mello et al. 2014).

Further, the primary use of Brazilian pasture land is low-intensity ranching of beef cattle, which can plausibly intensify with low or no cost. An appropriate suite of policies that achieves sugarcane ethanol expansion into pasture land, along with matching intensification in the beef industry, can assuage concerns about indirect land-use change (Cohn et al. 2014), as would the enforcement of restrictions on deforestation.

This chapter presents an optimization model of the Brazilian sugarcane industry that is used to simulate the effects of a range of assumptions about future variables. In particular, our main research question is: assuming free markets and a constant price, what would be the quantity and net present value (NPV) of the additional ethanol that firms produce in Brazil over the next 30 years. Further, we explore the impacts on this supply function of a global GHG price (or a policy that similarly affects the price of ethanol), aggressive technology investments resulting in high yield increases, evaluating the investment decisions in the model using a "social" discount rate, and different levels of available construction resources.

We set up and solve a mixed integer linear program (MILP) which allocates both new refineries and new sugarcane land over space and time to optimize total profits, depending on spatial variation in potential yield and freight cost to port and constrained by the available construction resources and pasture for conversion.

There are several novel features of this chapter as a contributor to the literature on the potential for renewable fuels and agricultural land-use change. We are the first to examine the supply of sugarcane ethanol in Brazil explicitly in an optimal investment framework. We use spatially disaggregated data on potential yield, freight costs, and available pasture land. We also explicitly account for the limits to investment over time due to constraints on construction capacity.

Most of the literature on the expansion of the biofuel sector analyzes the dynamics of feedstock area only (Khanna, Dhungana, and Clifton-Brown 2008). In contrast, our approach focuses on the investment in the processing facility and the change in the use of the land simultaneously. Consideration of the simultaneous adoption of several production technologies (i.e. a processing unit as well as feedstock units), as opposed to the adoption of an atomistic element of a supply chain, is particularly important for industries where a downstream subindustry depends on specific units of an upstream subindustry. In the case of sugarcane, the feedstock inputs are highly perishable and must be processed quickly, restricting the size of the catchment area for any given refinery. The sugarcane land would not be profitable without the nearby refinery, and vice versa. Essentially, the important industry feature is a high ratio of input to output freight costs, per unit value. This high ratio is a feature of almost every agricultural supply chain, in addition to many mining supply chains. The MILP framework is useful as it allows for the simultaneous modeling of the construction of the discrete refineries as well as the continuous feedstock land area around them.

In our primary results that assume recent oil-equivalent prices, our simulation model suggests that expansion of sugarcane ethanol over the coming decades would be unprofitable

in Brazil. However, as we add more optimistic assumptions, the outlook improves. In our most optimistic scenario, which assumes aggressive technology investments that result in high yield increases, evaluates investment decisions using a social discount rate, uses a GHG price, and allows for a large construction capacity, we calculate that 11% of global fuel liquids would be produced by Brazilian ethanol on 1.8% of global agricultural land, yielding $2859 billion in value. We believe that these more optimistic assumptions are safely in the space of reasonable parameters.

Background

Brazilians have both produced and consumed ethanol on a large scale in for many decades. In 2012, sugarcane for both ethanol and sugar production occurred on 9.8 Mha of land, and domestic production of ethanol amounted to 23.2Mm^3 (Brazilian Sugarcane Industry Association 2014). Lifecycle emissions of modern Brazilian sugarcane ethanol, absent land conversion, are just 13% of those associated with gasoline on an energy basis (State of California 2009). Mello et al. (2014) find that conversion from pasture to sugarcane in Brazil results in a cumulative reduction in soil organic carbon of 31.8Mg/ha over 20 years which, in our most optimistic scenario, is 5% of gasoline emissions, when averaged over 30 years of ethanol production. Estimates of emissions associated with indirect land-use change vary widely from -5 to 159 g CO_2-eq/MJ (Ahlgren and Di Lucia 2014), or -5.2% to 166% of gasoline emissions.

For a review of the numerous studies of global bioenergy potential, see Slade, Bauen, and Gross (2014). The majority of these studies do not examine any economic incentives; they just assess the physical potential of bioenergy production. To our knowledge, one article attempts to evaluate the global economic supply of biofuels; de Vries, van Vuuren, and Hoogwijk (2007) use the IMAGE model to estimate global bioenergy supply curves; however, the cost estimation, other than differences in yields, appears not to be spatially explicit beyond world regions (Hoogwijk 2004).

Two articles have performed more straightforward analyses of the physical potential of biofuel in Brazil in particular; Somerville et al. (2010) discuss the prospects for several biofuel crops, including Brazilian sugarcane, and Cerqueira Leite et al. (2009) ask what proportion of pasture land in Brazil would be required to displace 5% of world gasoline consumption. Both extrapolate existing yields to new sugarcane regions and perform straightforward multiplications of averages.[1] Accounting for the spatial variation in potential feedstock yields is crucial. Simple extrapolation based on empirical yields is inappropriate, as optimizing farmers will choose the land most appropriate for sugarcane first, so expansion areas will

[1]Somerville et al. arrive at a capacity of 500Mm^3 of conventional sugarcane ethanol per year in 2030 by multiplying the 2030 sugarcane yield prediction from FAPRI (96.47Mg Cane/ha) by their assumed area used for sugarcane ethanol production (59.28Mha) by the current typical ethanol conversion efficiency (0.086m^3 Ethanol/Mg Cane). Cerqueira Leite et al., in their main result, assume ethanol production of 102Mm^3, 71 Mg Cane/ha, and 0.085m^3 Ethanol/Mg Cane to arrive at 17Mha of land required for production.

likely be lower yielding than existing sugarcane land. For this reason, estimates of biomass potential that extrapolate using empirical yields will be biased upwards.

In the economics literature, Holland, Hughes, Knittel, and Parker (2014) employ a similar simulation model to ours in the US context. Their article examines the costs and benefits of various fuel policies and the correlation of these costs and benefits over space with voting on a cap-and-trade bill. Parker (2011) describes the underlying biofuel supply model. Like ours, it also solves a spatially explicit optimization problem designed to calculate the supply of biofuels; however, it does not account for the limited construction industry capacity to build new refineries, making it interpretable as a "long run" supply model. A model that does not account for supply over time can adequately answer questions relating to differences in profitability across space, as in Holland, Hughes, Knittel, and Parker (2014), but is not able to answer questions relating to the scale of potential production in the near to medium term. Considering the medium term potential is especially important in the biofuel supply context as it is not likely to become a dominant global energy source, due to the higher energy per unit land that solar photovoltaics can extract (e.g. Nelson (2010)).

This chapter proceeds as follows: the following section presents our high-level modeling framework, then describes our optimization model in detail. Next, we outline our full list of data sources then discuss our results. The final section concludes.

4.2 Conceptual framework

In this section, we describe the conceptual pieces that make up our larger simulation model. Firstly, to model the refinery investment decision, we use the standard NPV method, first formalized in Fisher (1907), where the refinery operator invests if the NPV is positive, and otherwise does not invest.[2]

To obtain indicative price changes for ethanol over time under a greenhouse price, we use a simple partial equilibrium model of the global energy market with constant elasticity demand and supply functions. The key assumptions we employ are: oil supply, ethanol supply, and energy demand are constant elasticity functions of price, oil and ethanol are perfect substitutes in the energy market, BTUs are the only valued component of either fuel, and oil BTUs are adjusted using a scalar multiplier to equate the prices of the two fuels on a per-BTU basis, before adjustments due to GHG pricing. We use the central parameter values from Holland, Hughes, Knittel, and Parker (2014). We fully describe the model in the Appendix.

[2]We do not account for uncertainty in this chapter. A more complicated, and realistic, investment decision rule that applies under uncertainty is optimal stopping (e.g. Dixit and Pindyck (1994)). However, in the usual way optimal stopping problems are set up, with just the output price uncertain, the decision rule is isomorphic to one where the operator invests if, and only if, NPV is above a threshold. One could then, in principle, map our analysis to one that uses optimal stopping investment rules and uncertainty, by identifying parameter values in the optimal stopping problem that produce the same marginal refinery. One could then interpret the prices that we assume in our calculations could as "certainty-equivalent" prices.

Both price expectations and constraints on refinery construction are not explicitly modeled but are of first order importance to our results. We assume prices are constant, before the effects of GHG pricing, and explore sensitivities to generate a full supply curve; in our specifications that calculate more outputs than just quantity, we assume a recently observed oil-equivalent price. We estimate the limit on refinery construction by doubling the historical maximum annual capacity installed, and we explore sensitivity to this assumption by tripling this limit.

Finally, we combine these elements in an optimization model, presented in full in the following subsection. Our model is similar in spirit to the classic models of von Thünen, or Alonso (1964), Mills (1967), and Muth (1969), with a monocentric destination for output and spatial allocation of firms driven by output freight costs and spatial variability in yields. A complete model would also expressly incorporate input freight costs and economies of scale in processing; firms would locate in resource-rich areas due to the trade-off between input and output freight costs, when the former are much larger than the latter per unit value, as is the case in our context. Economies of scale would then prevent processing units from being atomistic and thus driving input freight costs to zero. However, for computational tractability, we make simple restrictions to account for each of these features.

To account for input freight, we limit land to only come from nearby areas and assume a fixed cost per unit of sugarcane. To account for economies of scale in processing, we restrict all refineries built to have the same capacity, on the order of that for recently-built large refineries.[3] The amount of input land is then chosen to match this capacity.

Optimization model

In this section, we describe our simulation model, which we parameterize to the Brazilian sugarcane context. In this model, the unit of analysis is a municipality-year and all feedstock land comes from land that is currently in pasture. The model chooses the number of refineries built in each municipality-year and the amount of pasture purchased by each refinery operator.

The sugarcane cycle in Brazil is typically six years long, beginning with an initial planting that takes twelve to eighteen months to produce a first harvest, followed by four annual cuttings with declining productivity, finishing with a fallowing stage before the beginning of the next cycle. To account for this behavior, we assume the investor staggers her land purchases over six periods. At the time each production unit is built, the investor plants one sixth of the total area that will ultimately be used by the refinery. This land is first harvested in the following period.[4] The investor then continues to purchase and plant land

[3]The empirical relationship between capacity and construction cost is approximately linear, suggesting that the benefits from economies of scale lie outside the empirical range of refinery sizes.

[4]We do not model the dynamics of yields throughout the sugarcane lifecycle. This slightly biases the model against investing in ethanol as we force the model to transfer production from the early years of the cycle to the latter years.

for a total of six years. Five periods after the initial planting, the refinery operates at its capacity of 250ML/year, the size of a typical large ethanol refinery in Brazil. In the sixth period, the land that was first purchased is replanted, and the cycle restarts. Absent yield increases, this planting pattern allows the refinery to operate at a constant rate from the period five years following construction to the period 30 years following construction, after which we assume the project is abandoned and the land is sold at the initial purchase price. We account for yield increases in the model with simultaneous capacity upgrades, with costs proportional to the initial construction costs.

Refineries process all sugarcane into ethanol and send all production to Paulínia in São Paulo state, currently a delivery hub for both domestic consumption and export. We restrict the analysis to the Central-West, Southeast, and South regions, as well as Bahia state, to encapsulate all high-yielding areas that have pasture land available, to keep our freight destination assumption reasonable, and to maintain the tractability of the model.

We set up the supply model as a mixed integer linear program (MILP). For each run, we solve the problem to within 1% of the best objective bound using the optimization software Gurobi.

The full solution procedure is as follows. First, we calculate the NPV of revenue, per-hectare land operating costs, cane transport costs, refinery operating costs, construction costs, freight costs, and upgrade costs for a refinery located in each municipality i, and built at each time t. Next, we calculate the NPV of land conversion and purchase costs for a hectare located in each municipality i, and for each time t. For space, we omit the full definitions of these NPVs.[5] Finally, using Gurobi, we choose the (integer) number of refineries built in each municipality i and year t, and the (continuous) pasture land in municipality j used by refineries in each municipality i, where $j \in J_i$. J_i denotes i itself, i's direct spatial neighbors, and neighbors of neighbors.

The objective function is:

$$OBJ: \qquad \max_{K,H} \sum_{i=1}^{N} \sum_{t=0}^{T} (NPV_{NoLand,it} k_{it} - \sum_{j \in J_i} NPV_{Land,jt} h_{ijt}), \tag{4.1}$$

where $k_{it} \in K$ is the integer number of new refineries built in municipality i at time t, $h_{ijt} \in H$ is the amount of pasture (in hectares) used in municipality j by a refinery located in municipality i at time t. $NPV_{NoLand,it}$ is the NPV per refinery in municipality i built at time t, excluding land purchase and conversion costs. $NPV_{Land,jt}$ is the NPV of conversion and land costs per hectare in municipality j.[6]

[5] We provide all parameter values, with sources, in the Appendix. The calculation code is available from the authors on request.

[6] We do not account for emissions associated with land-use change in our simulation of the global oil and ethanol market, as these vary with the number of hectares, rather than the number of units of output. To account for this when we simulate the effects of a greenhouse gas price, we modify $NPV_{Land,jt}$ to include the change in soil organic carbon that results from the conversion from pasture to sugarcane from Mello

The first constraint ensures that the total pasture used does not exceed some maximum allowable limit $\bar{P}$. In all model runs presented in this chapter, we do not allow the total pasture converted to be more than 50% of what remains in Brazil; this acts as a constraint on the total eventual expansion of ethanol. In notation, this is:

$$\text{s.t.} \qquad \sum_{i=1}^{N} \sum_{t=0}^{T} \sum_{j \in J_i} h_{ijt} \leq \bar{P}. \tag{4.2}$$

Next, we ensure that the pasture land used in a municipality does not exceed the total amount in that municipality, which we denote $\bar{P}_i$:

$$\sum_{t=0}^{T} \sum_{j \in J_i} h_{jit} \leq \bar{P}_i \qquad \forall i, \tag{4.3}$$

where h_{jit} is the pasture used in i by a refinery in j at time t.

Recalling that refineries reach capacity after five years, the next constraint ensures that the sugarcane production from the land used by a refinery in municipality i, built at time t, matches the sugarcane production required by the refinery at time $t+5$ so that it operates at capacity:

$$\sum_{j \in J_i} h_{ijt} Y_{j,t+5} = k_{it} F \qquad \forall i, t, \tag{4.4}$$

where $Y_{j,t+5}$ is the yield of sugarcane per hectare of pasture in municipality j at time $t+5$, and F is the sugarcane required for a refinery to operate at capacity.

To substitute for explicit modeling of the supply curve of refinery construction, the final set of constraints limit the total number of refineries that can be built in each year. These constraints act as the main limit on the expansion of ethanol over time.

$$\sum_{i=1}^{N} k_{it} \leq K \qquad \forall t, \tag{4.5}$$

where K is the maximum refineries built in a year.

Because yields vary across municipalities, the total hectares required for a refinery can vary depending on the configuration of land chosen by the optimization algorithm. Further, because we assume linear yield trends, *along with matching increases in capacity*, the change in production of a unit over time (part of the NPV calculation) can also depend

et al. (2014), prorated over the first 20 years from conversion, and monetized using the same GHG prices as above.

on this configuration, which is chosen by the MILP. Thus, the correct $NPV_{NoLand,it}$ coefficients in Equation 4.1 are determined by the final model solution, making the optimization problem nonlinear. To overcome this issue, we simplify by assuming that, when calculating $NPV_{NoLand,it}$, the production increases for a refinery in municipality i are the same as those that would result from a refinery that only used land from within municipality i.[7] If a refinery were to use a neighboring municipality's land, which was lower yielding than the land in its own municipality, this assumption would have the effect of lowering production increases over time, as the number of hectares per refinery would be lower than what the optimization chooses.

4.3 Data

This section describes the input data for our model, along with sources. Our general aim is to incorporate all readily available information and to be conservative when making assumptions; that is, biased against investment in ethanol.

Yield data and calculations

The FAO-GAEZ dataset (IIASA and FAO 2012) uses agronomic modeling, based on climate, soil and terrain data, to provide global gridded datasets of potential yields for many important crops, including sugarcane. These data give the relative spatial arrangement of the yield assumptions in our analysis. All scenarios presented in this chapter use the yield values predicted for the 2020s using the Hadley CM3 A1F1 scenario,[8] assuming "high" input levels[9] and rain as the water supply.[10] The grid is then aggregated up to the municipality-level on an area-weighted basis. Figure 4.1 displays the FAO-GAEZ yields for the areas in our analysis.

We obtain empirical yields for São Paulo from the Brazilian Sugarcane Industry Association (UNICA), which provides a long time series of state-level yields. We collect municipality-level planted areas for 2012 from the Brazilian Institute of Geography and Statistics (IBGE).

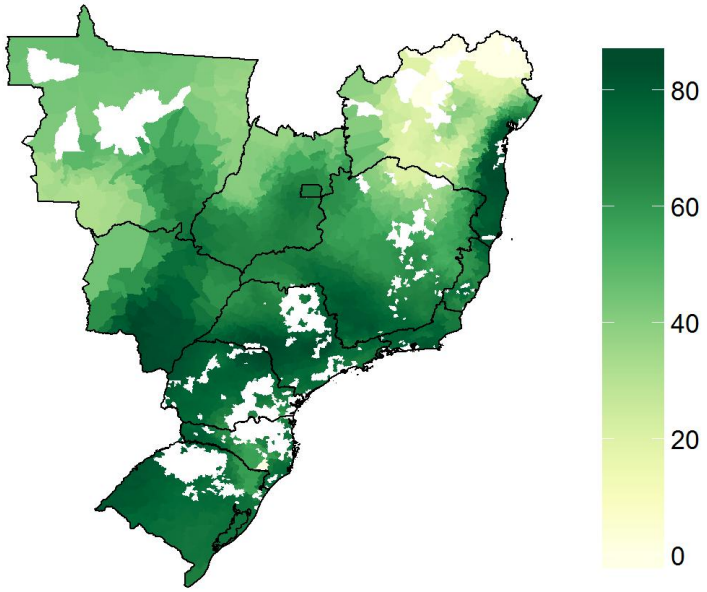
[7] The spatial autocorrelation of rain-fed potential yields for neighbors within two spatial lags is 0.91.

[8] The SRES A1F1 scenario corresponds closely in predicted temperature increases to the RCP8.5 scenario (Rogelj, Meinshausen, and Knutti 2012).

[9] The FAO-GAEZ documentation states: "Under a high level of input (advanced management assumption), the farming system is mainly market-oriented. Commercial production is a management objective. Production is based on improved or high yielding varieties, is fully mechanized with low labor intensity and uses optimum applications of nutrients and chemical pest, disease and weed control" (IIASA and FAO 2012, p. 38).

[10] We do not allow our model to invest in irrigation as we were unable to find reliable cost estimates for irrigation installation. The omission of irrigation as an option biases our results against the construction of new ethanol capacity.

Figure 4.1: Municipality-level potential yield predictions



Obtained from IIASA and FAO (2012). Results displayed for the Central-West, Southeast, and South regions, as well as Bahia state. Predictions are for the 2020s using the Hadley CM3 A1F1 scenario, and rain-fed systems. Omitted municipalities have either zero potential yield or zero pasture land available.

Potential yields derived from the FAO-GAEZ data are somewhat higher than empirical yields in the regions of Brazil with already developed sugarcane land. To correct for this, we scale down the FAO-GAEZ yields so that the planted-area-weighted São Paulo (the most advanced sugarcane area) empirical average yield matches the FAO-GAEZ average for the same region. This process ensures that the yield assumptions are initially at status quo levels and that the arrangement of these across space reflects variation in climate, soil, and terrain.

The empirical data only reports the average of rain-fed and irrigated yields, whereas the FAO-GAEZ data separates these. We recover separate estimates by assuming that the true ratio is equal to the ratio between the irrigated and rain-fed yields in the FAO-GAEZ dataset. That is, we solve the following equations for $\bar{y}_R$ and $\bar{y}_I$:

$$\begin{aligned} \bar{y} &= \rho_I \bar{y}_I + (1 - \rho_I)\bar{y}_R \\ \frac{\bar{y}_R}{\bar{y}_I} &= \frac{\bar{y}_{R-GAEZ}}{\bar{y}_{I-GAEZ}}, \end{aligned}$$

where $\bar{y}$ is the empirical planted-area weighted sugarcane yield, taken from the UNICA data, $\rho_I = 0.39$ is the proportion of sugarcane land irrigated,[11] $\bar{y}_R$ is the empirical planted-area

[11]We know of no detailed data on sugarcane irrigation levels in Brazil. We use the national average

weighted sugarcane yield for rain-fed plots, $\bar{y}_I$ is the empirical planted-area weighted sugarcane yield for irrigated plots, $\bar{y}_{R-GAEZ}$ is the planted-area weighted FAO-GAEZ predicted potential yield for rain-fed plots, and $\bar{y}_{I-GAEZ}$ is the planted-area weighted FAO-GAEZ predicted potential yield for irrigated plots. The two unknowns in the above are $\bar{y}_R$ and $\bar{y}_I$ and all yield averages here are for São Paulo. Municipality level planted areas, used for aggregating the GAEZ data up to the state level, are taken from the IBGE data.

Next, we scale the municipality-level FAO-GAEZ yields for all municipalities to get our first-period yield assumptions:

$$y_{iR0,Final} = \frac{\bar{y}_{R0}^{SP}}{\bar{y}_{R-GAEZ}^{SP}} * y_{iR,GAEZ}$$

where i indexes municipalities, $\bar{y}_{R0}^{SP}$ denotes the empirical yield in São Paulo in our first model year.

Finally, to incorporate expected productivity increases, we estimate the yield trend using ordinary least squares (OLS) for São Paulo, using the UNICA data, and predict for all years in our analysis (2014–2068).[12] Because São Paulo is the Brazilian state with the most developed sugarcane industry, these figures represent a reasonable point prediction of ongoing yield increases, given normal to good production practices.

$$y_{iRt,Final} = y_{iR0,Final} + 0.3243t$$

In an alternative scenario, we also assume high yield increases, that could eventuate from more aggressive investment in yield-improving technology, of 2 Mg/ha. Near the end of our study period, some areas' assumed yields would then be close to recent experimental maxima reported in Waclawovsky et al. (2010).[13]

Land Prices

Because Brazilian land price survey data is proprietary, we simply use a recent academic article which reports averages for 2002, 2006, and 2010 (Richards, Walker, and Arima 2014). Because the values have trended slightly downwards in real terms, instead of extending the trend, we conservatively assume pasture prices average to the most recent value, 2014R$1514/ha. We then assume land prices can be expressed as a linear function of the period 0 rain-fed yield as calculated in the previous subsection. Because land prices comprise a

proportion to proxy for the São Paulo proportion (Soybean And Corn Advisor 2014; Brazilian Sugarcane Industry Association 2014).

[12]We use data for 1980-2012 and outliers near the beginning and end of the sample period (1981, 1982, 1983, 2000, 2011, and 2012). Keeping the outliers produces a prediction around 0.8% higher than that which is used. $\bar{y}_{R0}^{SP}$ is calculated in this prediction.

[13]Note that the maximum yield that we assume over *all* scenarios and time periods we consider is 55% of the *theoretical* maximum reported in Waclawovsky et al. (2010).

small fraction of total costs in our model, the qualitative results are insensitive to reasonable adjustments to these assumptions. Importantly, the prices we use were originally derived from sales data, which is indicative of land values conditional on the land being sold. If sold land is systematically higher or lower value than average, our land price assumptions may be biased. However, as aforementioned, because land prices comprise such a small fraction of total costs, this bias would have to be large to affect our results materially.

In addition to the land purchase price, we obtain land conversion costs from Bonomi et al. (2012). These costs cover roads construction, terraces construction, agricultural area systematization, and roads maintenance. Again, these costs are small in comparison to total costs, so results are insensitive to reasonable adjustments to this input.

Refinery construction, upgrade costs, and construction industry capacity

Refinery construction costs are estimated using data obtained from the Bloomberg New Energy Finance (BNEF) database. This database includes construction costs and capacities of many ethanol refineries built in Brazil. As aforementioned, we simplify our model by restricting each refinery to have a capacity of 250ML/year. We estimate the cost of each of these plants by regressing real construction costs on capacity and a trend line, and predicting for 250ML/year and our first model period[14]. Refineries included in this estimation are those that are ethanol-only, located in Brazil, and use sugar crops as the feedstock. Upgrade costs are assumed to be equal to construction costs on a per-unit-capacity basis.

The same dataset is used to obtain a reasonable assumption of how many refineries might be able to be constructed during any year. We assume that the industry can double the capacity increase of the highest construction year, which was 4548ML. We observe a total of 3735.6ML nameplate capacity for 23 refineries built in 2009 in the BNEF data, of 28 total (Barros 2014). We estimate total new capacity in 2009 using the product of the total capacity observed in the BNEF data and the ratio of total refineries built to the number observed in the BNEF data.[15]

Operating Costs

We obtain typical annual operating costs from both PECEGE (2012) and Bonomi et al. (2012). In the model, we collapse these costs into a per-hectare component, which does not increase with yields, a per-unit of sugarcane component, and a per-unit of ethanol component.

The per-hectare component covers all expenses after land conversion costs and before transportation costs from the field to the refinery, not including rent. These include costs

[14]The trend is slightly downward in real terms, but statistically insignificant.

[15]This is 28/23*3735.6 = 4548.

related to planting, fertilizer, and harvesting. The number is obtained by taking total operating expenses less land rent and sugarcane transportation, for refinery operated sugarcane operations in the expansion region, from PECEGE (2012).[16]

The per-unit-of-sugarcane component accounts for the cost of sugarcane transport from field to refinery per metric ton of cane from Bonomi et al. (2012).

The per-unit-of-ethanol component captures all refinery processing and maintenance costs. We calculate this as the total cost less feedstock cost, depreciation, cost of capital, and working capital.

In all specifications, we keep these operating costs constant in real terms.

Ethanol Freight Costs

For simplicity, we assume all ethanol freight goes to Paulínia, SP, the delivery destination for the BM&F Bovespa ethanol futures contract, and the location of a major delivery hub.

We obtain empirical intercity ethanol freight costs per m^3 from ESALQ-LOG (2013). Because the municipalities we allow in our analysis are far more numerous than those observed in the empirical data, we use a simple predictive model for freight costs, based on distance by road. The Google Maps API is used to obtain road distances between all origin-destination pairs in ESALQ-LOG (2013). We then estimate the following predictive relationship using OLS:[17]

$$F_{ij} = \alpha + \beta_1 D_{ij} + \beta_2 D_{ij}^2 \tag{4.6}$$

Finally, we use the Google Maps API to obtain the road distance to Paulínia, SP for all municipalities and predict the freight costs using (4.6).

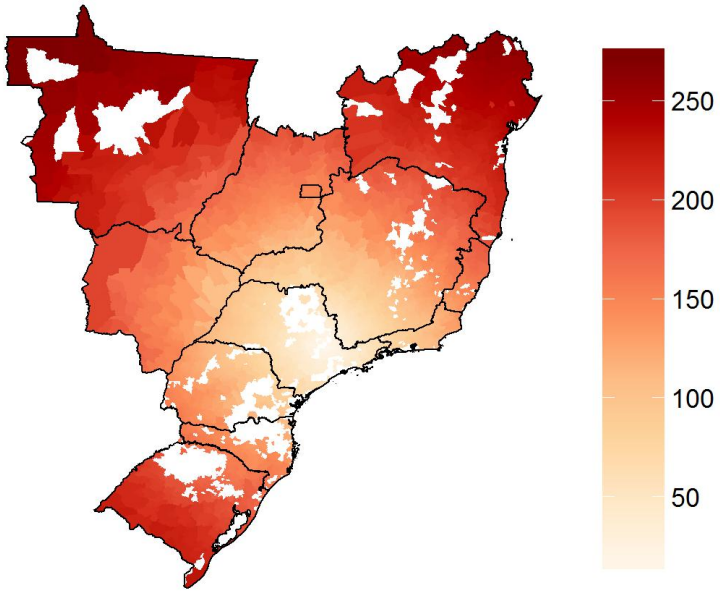
Pasture Location Data

A shapefile containing pasture data is obtained from IBGE (2013). This file contains the locations of pasture farms in Brazil as of 2012. We calculate the pasture land available in each municipality by spatially aggregating this shapefile up to the municipality-level and computing the total area.

[16]Total operating costs and land rent are collected from table 14 in PECEGE (2012). The operating expenses category includes costs of transport from the field to the refinery, but it is not specifically itemized. To calculate the adjustment, we use table 6 in Bonomi et al. (2012) and calculate the proportion that transportation contributes to costs related to planting, cultivation, harvesting, and transportation, and subtract this from the operating expenses in PECEGE (2012).

[17]We trialed several alternative polynomial specifications, including ones that include elevation as a predictor. None substantially improve the predictive power of the estimated relationship.

Figure 4.2: Model predicted freight rates per m^3 of ethanol



Raw data obtained from ESALQ-LOG (2013). Unobserved routes are extrapolated using a quadratic predictive relationship between freight rates and roading distances.

Output Prices

In our results, we calculate supply curves over many assumed output prices. For comparison to status quo prices, we collect output prices for ethanol from BM&F Bovespa (2015). We use a recently collected futures prices for hydrous ethanol delivered in São Paulo (delivery October, 2016, collected October 25, 2016).[18].

Because ethanol has traded at a premium over oil on an energy basis, we also compare our supply curves to an oil-equivalent price. The ethanol BTU premium/discount is due to a number of factors, including oil refining costs, the value of ethanol as an oxygenate/octane enhancer, the lower energy density of ethanol (meaning transport and storage costs are higher for ethanol), short run market conditions, and relative policy support for ethanol versus gasoline. As of October 25, 2016, the ethanol premium, adjusted only for the lower energy density of ethanol, is 209%. After accounting for refining costs, calculated using the United States Energy Information Administration's (EIA) decomposition of the gasoline price for September, 2016, the premium becomes 123%. When accounting for a further USD$0.25/gallon value to account for ethanol's value as an oxygenate (as suggested in Hurt, Tyner, and Doering (2006)), ethanol has an implied premium of 98%. When we explore model outputs in addition to quantity, we assume an oil-equivalent price that removes the

[18]There are few ethanol futures contracts for Brazilian delivery. As of October 25, 2016, the contract with the most distant delivery was March, 2017.

98% premium.

To generate indicative price changes that may result from a global greenhouse price, we simulate a simple global oil and ethanol market with GHG pricing, discussed further in the Appendix. Oil prices for this exercise, and for calculating the above oil-equivalent price, are collected from CME Group (2015), and we use all available futures prices that do not appear stale.[19]

Greenhouse Gases

In the aforementioned oil and ethanol market exercise, we use the values for the social cost of carbon (SCC) emitted in different years, as calculated by the Interagency Working Group on Social Cost of Carbon of the United States Government (US EPA Climate Change Division 2013), to calculate a total social value of the GHG emissions of oil and ethanol respectively.[20] We use the values reported using a 3% discount rate and linearly interpolate between years.

We obtain the lifecycle GHG emissions of sugarcane ethanol from California's Low Carbon Fuel Standard (LCFS) documentation. We make use of the estimate that does not take account of indirect land-use emissions. The LCFS implicitly assumes that carbon fluxes that directly arise from changes in land-use are zero, so we also account for changes in soil organic carbon due to the permanent conversion of land from pasture to sugarcane using Mello et al. (2014).

We obtain lifecycle GHG emissions of conventional oil from Chavez-Rodriguez and Nebra (2010).

4.4 Results

We present results for six scenarios. The first is the "Reference" scenario, where we assume free and efficient land markets, constant prices, and otherwise make assumptions that reflect the status quo.[21] Second, we present a "GHG price" scenario that features an increasing profile of global GHG prices over time (or equivalent domestic policy), resulting in immediately higher and further increasing ethanol prices over time. Third, we show a "Social discount rate" scenario where we assume a 3% real discount rate, versus the reference 6.08%,[22] which some argue is more appropriate than financial investment rates when evaluating the impacts of government policies. Fourth, in our "High yield increase" scenario, we assume aggressive investments in sugarcane technology that produce an annual yield increase of 2 Mg, versus

[19] We define this to be all monthly prices before the first change of more than 5% between consecutive months.

[20] See Greenstone, Kopits, and Wolverton (2013) for a discussion of the methodology of this process.

[21] The full set of parameter values, with sources, is presented in the Appendix.

[22] The "Reference" scenario discount rate is calculated using the nominal USD weighted average cost of capital for Petrobras, obtained from www.wikiwealth.com, adjusted for expected USD inflation, obtained from www.tradingeconomics.com.

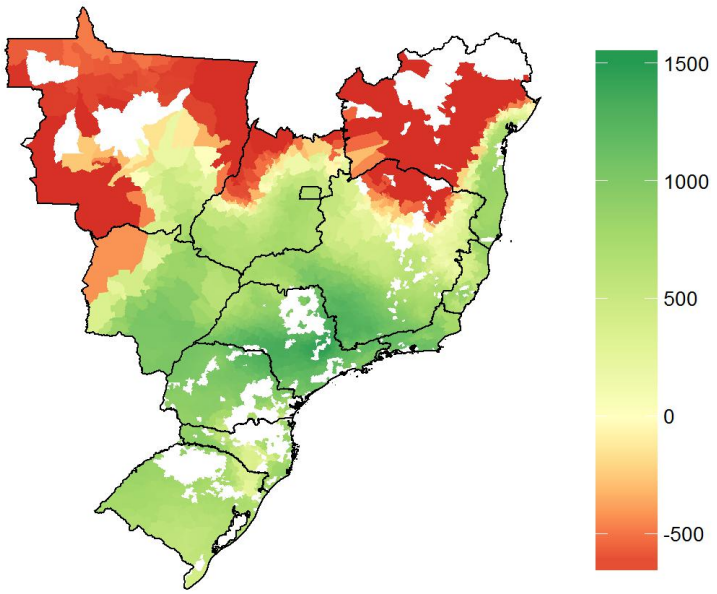
the reference 0.32 Mg. Fifth, our "High construction capacity" scenario substantially relaxes the constraint on building by allowing 108 refineries to be built per year, versus the reference 36.[23] Finally, in our "All" scenario, we assume all the changes in the "GHG price", "Social discount rate", "High yield increase", and "High construction capacity" scenarios.

Profitability over space

First, we can see that in the "Reference" scenario, the model predicts no expansion of the ethanol sector in Brazil under the subsidy-free price. We also see that the model predicts such large production growth that the refinery building constraint binds for *all years* at the current ethanol price. This price increase is reasonably large (98%); however, only an 18% increase is required for the model to predict enough expansion to meet the construction capacity constraint for all periods. The modest slope is mostly driven by the small variation in yields and freight costs over the space that's initially invested in (i.e. near Paulínia), and the low pasture land prices we see in Brazil. As aforementioned, because our cost model includes a substantial per-hectare cost component, the largest contributor to the spatial variation in profitability is differences in yields.

[23] In this scenario, the annual increase in capacity is 27000ML. For comparison, the maximum historical annual production increase in the USA was 10554ML in 2008.

Figure 4.3: NPV per refinery in the "GHG price" scenario



Values are for refineries constructed at time 0 and are censored at the negative of the refinery construction cost.

In the "GHG price" scenario, we plot the ethanol price *before* adjustments due to GHG pricing on the vertical axis. Recall that the GHG price, and thus the ethanol price, is increasing over time in these scenarios. The ethanol price faced by the refineries in these scenarios is substantially higher than in the "Reference" scenario, resulting in investment taking place at much lower initial ethanol prices. Here, the supply curve shifts down and becomes less elastic.[24]

Assuming aggressive investments in technology that result in high yield increases generates both a supply curve that is somewhat less elastic, and extends to much higher quantities of total ethanol production when the building constraint binds, as each refinery eventually upgrades by a larger amount, when compared to the previous scenarios.

Upweighting more distant cash flows by utilizing a social discount rate likewise shifts the curve down. However, the relative importance of this variable, when compared to GHG pricing, is small.

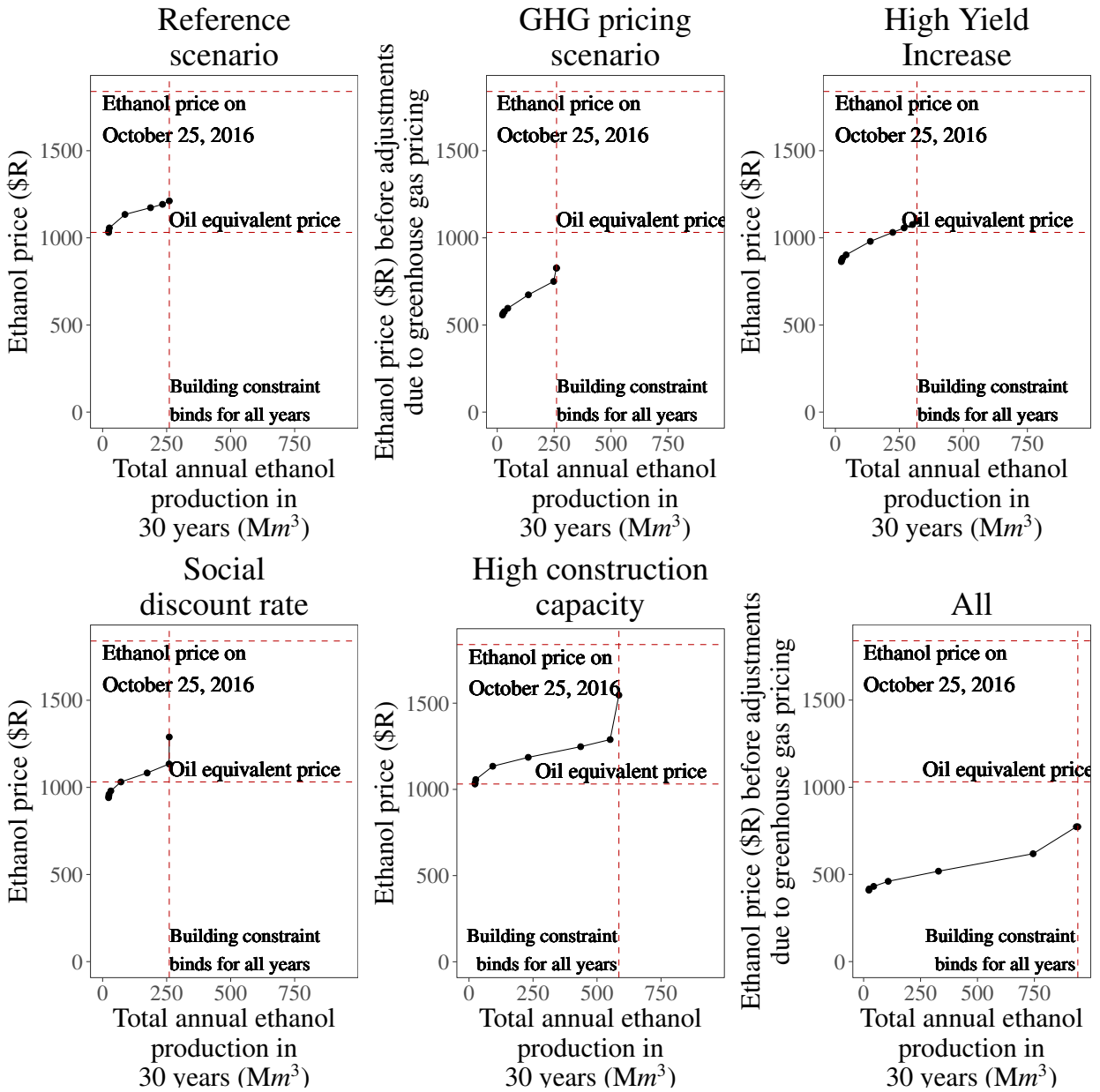
The "High construction capacity" scenario highlights the importance of the refinery building constraint. While not substantially changing the supply curve in the region of total production where the limits on refinery construction do not bind for all periods, relaxing this constraint increases the total potential production considerably.

In our "All" scenario, which combines the adjustments from all of the previous four scenarios, we see a combination of the shifting down and extending the supply curve. What is striking here is the scale of production implied at today's oil-equivalent price. The model suggests that Brazil can eventually economically produce levels of sugarcane ethanol energy at a similar level of that which Saudi Arabia and Russia produce in oil energy today.

Because we find such small slopes on the supply curves, the construction capacity constraint binds at many different reasonable prices, even in the scenarios that partially relax this. An improvement to this modeling exercise would be to directly model the construction supply curve to reflect increasing scarcity of human and physical capital inputs into this process. While beyond the scope of the current project, this improvement is an important avenue for future research that would move further towards a complete characterization of the supply function for ethanol.

[24]The slope increase is driven by the result that, under GHG pricing, the most profitable investments occur in the later periods, because the combined effect of the increasing price profile over time and yield increases outweighs the effect of discounting. So, the marginal refinery can now be in an earlier period, where the initial price facing the refinery is now lower (due to the increasing price profile). So, the price increase required to bring the marginal refinery into production must be higher, when compared to the "Reference" scenario.

Figure 4.4: Calculated supply curves for Brazilian ethanol production in 30 years



Total production is the sum of existing ethanol production and new model-predicted ethanol production. World crude oil supply in 30 years is projected to be energy equivalent to 5765 Mm^3 of ethanol. Optimization is performed to within a 1.5% MILP gap to reduce the computational burden. Oil equivalent price removes the ethanol BTU premium but retains the premiums associated with ethanol's value as an oxygenate and oil's need for further costly refining.

Various model outputs

For each scenario, table 4.1 presents each of total NPV, investment cost (excluding land purchases), and the amount of pasture converted, again using the oil-equivalent price. Before the discussion, note well that the analysis in our model abstracts from both taxation and the increases in land prices that would likely result from the prospect of substantial investment in the sugarcane industry. As such, the NPV numbers should be interpreted as the amounts to be shared between refinery/sugarcane investors, government revenue, and current landowners.

In our "Reference" and "High construction capacity" scenarios, the model finds that no new refineries are profitable. Using a social discount rate or assuming fast-growing yields increases the profitability somewhat, so that the model predicts aggregate NPVs of $4.4 and $18 billion for these scenarios respectively. These both represent substantial expansions over current production.

Introducing a GHG price greatly increases the profitability of sugarcane ethanol to yield a model-predicted NPV of $272 billion. Production, in this scenario, is limited by the refinery construction constraint.

However, the most interesting result from this table is that the combination of all the optimistic deviations from the "Reference" scenario results in investments totaling almost $3 trillion in present value. This massive increase in profitability, when compared with the

Table 4.1: Aggregate profit, investment cost, and land used by model scenario

Scenario	Total NPV ($ billion)	Investment Cost NPV (Land Conversion and Refinery) ($ billion)	Pasture Converted (Mha)
Reference scenario	0	0	0
GHG price	271.7	115.7	31.51
Social discount rate (3%)	4.411	23.35	6.178
High yield increase (2 tonne/year)	17.52	74.04	16.35
High construction capacity (108 refineries/year)	0	0	0
All	2859	672.6	84.4

Total NPV is the sum of the NPV of refinery construction costs, land purchase costs, land conversion costs, feedstock costs (production and transport), refinery operating costs, and ethanol freight costs. Investment Cost includes land conversion, refinery construction costs, and refinery upgrade costs. There is approximately 4,912 Mha of agricultural land and 3,359 Mha of pasture land globally. There is approximately 281 Mha of agricultural land and 172 Mha of pasture land in Brazil.

Table 4.2: Decomposition of NPV by scenario

Scenario	Revenue	Refinery Operation	Land Operation	Construction/ Upgrade	Freight	Other
Reference/High construction capacity	0.00	0.00	0.00	0.00	0.00	0.00
GHG price	824.97	131.78	211.62	109.40	42.21	58.23
Social discount rate (3%)	152.23	41.10	60.58	23.17	7.69	15.28
High yield increase (2 tonne/year)	319.95	86.39	88.01	73.69	21.58	32.76
All	6588.36	1015.09	1121.54	641.63	557.47	393.45

Revenue accounts for all receipts at the delivery point. Refinery operation accounts for all operating expenses at the refinery. Land operation accounts for all operating expenses in sugarcane fields, including capital depreciation, and excluding rent. Construction/Upgrade accounts for all capital expenditure at the refinery. Freight accounts for transportation costs from the refinery to the delivery point. Other includes the cost of transporting sugarcane from the field to the refinery, land purchases, and land conversions.

Table 4.3: Aggregate output and percentage of global agricultural area by model scenario

Scenario	Percentage of Global Agricultural Area	Total Production in 30 Years (million m^3/year)	Total Brazil Ethanol as Percentage of World Liquids in 30 Years
Reference scenario	0.1%	26.66	0.3%
GHG price	0.74%	264.2	2.9%
Social discount rate (3%)	0.23%	74.8	0.83%
High yield increase (2 tonne/year)	0.43%	245.1	2.7%
High construction capacity (108 refineries/year)	0.1%	26.66	0.3%
All	1.8%	997.1	11%

Total production includes both current production and the model-predicted new production. 30 year projection of world liquids is linearly extrapolated from the BP world energy outlook. World crude oil supply in 30 years is projected to be energy-equivalent to 5765 Mm^3 of ethanol.

"GHG price" scenario, arises from the combination of several complementary effects. First, much more production occurs due to a tripling of the allowed number of refineries built in the model. Next, large increases in both output and profit per unit arise due to high yield growth; the latter effect is occurring as we have a large per-hectare cost component. Lastly, future cash flows are up-weighted when using the social discount rate. Table 4.2 decomposes the NPV by scenario into several categories, allowing the reader to further explore how the components change by scenario.

Table 4.3 presents two measures of aggregate production predicted by the model in each of the scenarios; these are total output in 30 years, both expressed as a level and as a percentage of world liquids production in 30 years, and the proportion of land employed in sugarcane in Brazil.

In the "Reference" and "High construction capacity" scenarios, no new production occurs so we report only existing production, which we assume will continue as is. In the "Social discount rate" scenario, many areas become profitable, and investments are made to increase production to almost 1% of world liquids production.

When we assume high yield increases, many more areas become profitable and, because those areas are also more productive, total output increases to around 3% of global liquids production, an amount similar to that of a top 10 oil producing country today. The "GHG price" scenario, where the refinery construction constraint binds, yields a similar level of production on a larger amount of land.

Again, it is the optimistic scenario that contains the most striking result here. When allowing for highly profitable production through a GHG price, a large amount of construction resources, and high yield increases, we calculate that Brazil would produce more liquids energy than the USA does today. When adding Brazil's current oil production, this would make it the largest liquids producer globally.

Limitations

Our model predicts substantial investment *today*, at current ethanol prices. There are several real-world barriers to investment that can account for this disconnect from the relatively low levels of sugarcane ethanol investment we have observed empirically in recent years.

Firstly, due to the mandate, the current ethanol price includes an observed premium over oil, even when adjusting for refining costs and ethanol's value as an oxygenate. So, the current ethanol price does not reflect the prevailing energy price. However, the implicit subsidy in Brazil is highly unlikely to be reduced substantially in the near term, so real-world price expectations should likely be formed with this premium for several years to come.

Capital controls and general uncertainty over the stability of institutions in Brazil can partially account for low investment. These capital controls manifest in restrictions on the amount of land that can be controlled (including leasing) by foreign interests. However, our discussions with local experts indicate that no such restrictions exist on contracts with local farmers, so presumably the development of this institutional arrangement, in this con-

text, could be a path forward for would-be investors. However, it is also plausible that the government could view long-term production contracts with local landholders as a form of leasing. The magnitude of the effect of general institutional instability cannot be known, but it surely non-zero.

We do not model yield variability, which could reduce capacity utilization below what it is in the model. We do not model any land market frictions, essentially assuming eminent domain. Uncertainty over sugarcane production in the refinery catchment area would reduce investment.

Our model makes the simplifying assumption that real ethanol prices will remain the same over the investment period. As ethanol production expands, obviously demand will also have to increase to keep prices constant. While this is simply an assumed scenario, it is important for the reader to understand the type of world this imagines. For example, a scenario in which ethanol blending and E85 are progressively adopted globally would be consistent with our more optimistic results, as we project up to 11% of global energy liquids supply will come from Brazilian biofuel.

We know of no study that examines the effect of price volatility on investment in ethanol refineries specifically, through the mechanism of utilizing the option value of investment delay (Dixit and Pindyck 1994). Kellogg (2014) uses data on Texas drilling operations to indicate the effect of uncertainty on investment in an empirical context. Extending Kellogg's result to the full certainty case suggests that uncertainty can account for approximately a 25% reduction in investment rates, in that circumstance. However, the effect of uncertainty in any given context is highly dependent on the level of profitability of the investment, so a parametrization for the Brazilian ethanol context would be required to better understand how large the magnitude of this effect is.[25] This would be a fruitful avenue for future research.

We also do not directly model the supply curve for refinery construction, choosing to make the simplifying assumption that the number of refineries built is limited to 36 per year in most scenarios and 108 in the "Low Building Constraint" and "All" scenarios. There are several reasons why direct modeling of the refinery construction supply curve is difficult in our context. Firstly, an increasing construction cost curve would make our optimization model nonlinear, increasing an already large computational burden. Secondly, even simple empirical estimation of this supply curve is limited by the few observations of ethanol refineries constructed in Brazil. Thirdly, careful modeling of the refinery construction process is outside the scope of this chapter. Doing so would be another avenue for future research.

Another potentially important omission is any modeling of the sugar market. However, because this chapter is primarily focused on scenarios in which sugarcane production for ethanol is vastly expanded, the relative importance of the sugar market will be much diminished.

There are also several limitations of our model that may bias us against investment in ethanol. Of first order concern is the relative future cost reductions in ethanol versus other

[25] See figure 6 in Kellogg (2014, pp. 1715).

transportation fuels. For example, plant growth technologies, such as CRISPR (Doudna and Charpentier 2014), could potentially vastly reduce the cost of producing sugarcane, and decrease conversion costs if sugar density increases. Cost reductions in second generation biofuels could similarly vastly improve the profitability of ethanol. Of course, these cost reductions have to outpace reductions in the costs of production of other fuels to bias our results against investment in ethanol.

We also do not account for income associated with the sale of electricity from burning bagasse, which could add around 43% to revenues in optimized refineries.[26] This, potentially significant, omission likely biases our results away from investment in ethanol but is difficult to model as we can't separately observe the construction costs of the electricity production infrastructure in the refineries, and a full analysis would need to account for the potentially large transmission costs to bring the electricity to population centers. More detailed modeling of the use of bagasse in this context will also be a fruitful area for future research.

Finally, restricting refinery size to be fixed removes a dimension of optimization, also biasing the model against ethanol investment. We also do not allow for any second generation ethanol production, so if this technology becomes economic, our model would underestimate refinery values.

4.5 Conclusion

This chapter develops a supply model for ethanol production intending to assess the economic potential of biofuel in Brazil under a variety of future scenarios. We show that, with free capital markets, constant prices, and a GHG price, a non-trivial amount of future global liquid fossil fuel can be profitably displaced by ethanol production using existing pasture land. Because the GHG price increases profitability by so much, our model predicts that incorporating high yield increases and a large capacity for constructing refineries would increase production further so that 11% of global liquids production would come from Brazilian ethanol, using 1.8% of global agricultural land.

[26] This is calculated using the ratio of surplus electricity to ethanol produced per tonne of cane in figure 16 of Bonomi et al. (2012, pp. 69) in kWh/L, multiplied by 1000/1000 to "change" the units to MWh/m^3, multiplied by R$200, which is a reasonable average wholesale spot price in Brazil, and divided by 1031, which is our oil-equivalent price.

Bibliography

Ahlgren, Serina and Lorenzo Di Lucia (2014). "Indirect Land Use Changes of Biofuel Production – a Review of Modelling Efforts and Policy Developments in the European Union". In: *Biotechnology for Biofuels* 7, p. 35. DOI: 10.1186/1754-6834-7-35.

Aigner, D. J. (1974). "MSE Dominance of Least Squares with Errors-of-Observation". In: *Journal of Econometrics* 2.4, pp. 365–372. DOI: 10.1016/0304-4076(74)90020-7.

Alonso, William (1964). *Location and Land Use*. Cambridge, MA: Harvard University Press. URL: http://www.hup.harvard.edu/catalog.php?isbn=9780674730854.

Andreou, Elena, Eric Ghysels, and Andros Kourtellos (2010). "Regression Models with Mixed Sampling Frequencies". In: *Journal of Econometrics*. Specification Analysis in Honor of Phoebus J. Dhrymes 158.2, pp. 246–261. DOI: 10.1016/j.jeconom.2010.01.004.

Atkinson, Scott E and Thomas D Crocker (1992). "Econometric Health Production Functions: Relative Bias from Omitted Variables and Measurement Error". In: *Journal of Environmental Economics and Management* 22.1, pp. 12–24. DOI: 10.1016/0095-0696(92)90016-P.

Auffhammer, Maximilian, Solomon M. Hsiang, et al. (2013). "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change". In: *Review of Environmental Economics and Policy* 7.2, pp. 181–198. DOI: 10.1093/reep/ret016.

Auffhammer, Maximilian and Wolfram Schlenker (2014). "Empirical Studies on Agricultural Impacts and Adaptation". In: *Energy Economics* 46, pp. 555–561. DOI: 10.1016/j.eneco.2014.09.010.

Barros, Sergio (2014). *Brazil Biofuels Annual*. BR14004. Washington DC: USDA.

Bell, Kendon M. (2017). "Estimation of the Costs and Benefits of Climate Change Mitigation". Ph.D Dissertation. University of California, Berkeley. 98 pp.

Blanc, Elodie and Benjamin Sultan (2015). "Emulating Maize Yields from Global Gridded Crop Models Using Statistical Estimates". In: *Agricultural and Forest Meteorology* 214–215, pp. 134–147. DOI: 10.1016/j.agrformet.2015.08.256.

Blomqvist, A. G. (1972). "Approximating the Least-Squares Bias in Multiple Regression with Errors in Variables". In: *The Review of Economics and Statistics* 54.2, pp. 202–204. DOI: 10.2307/1926284.

BM&F Bovespa (2015). *Settlement Prices*. URL: http://www2.bmf.com.br/pages/portal/bmfbovespa/boletim2/Ajustes2.asp (visited on 01/09/2015).

Bonomi, Antonio et al. (2012). *Technological Assessment Program (PAT) The Virtual Sugarcane Biorefinery (VSB) 2011 Report*. Technical Report. Campinas, Brazil: Brazilian Bioethanol Science and Technology Laboratory (CTBE).

Brazilian Institute for Geography and Statistics (2013). *Technical Manual on Land Use [Portuguese: Manual Técnico de Uso Da Terra]*. Rio de Janeiro: Brazilian Institute for Geography and Statistics.

Brazilian Securities, Commodities and Futures Exchange (2014). *Settlement Prices*. URL: http://www2.bmf.com.br/pages/portal/bmfbovespa/boletim2/Ajustes2.asp (visited on 01/05/2015).

Brazilian Sugarcane Industry Association (2014). *Unicadata*. URL: http://www.unicadata.com.br/ (visited on 10/08/2014).

Burke, Marshall B., John Dykema, et al. (2014). "Incorporating Climate Uncertainty into Estimates of Climate Change Impacts". In: *Review of Economics and Statistics* 97.2, pp. 461–471. DOI: 10.1162/REST_a_00478.

Burke, Marshall B. and Kyle Emerick (2016). "Adaptation to Climate Change: Evidence from US Agriculture". In: *American Economic Journal: Economic Policy* 8.3, pp. 106–140. DOI: 10.1257/pol.20130025.

Burke, Marshall B., Solomon M. Hsiang, and Edward Miguel (2015). "Global Non-Linear Effect of Temperature on Economic Production". In: *Nature* 527.7577, pp. 235–239. DOI: 10.1038/nature15725.

Burke, Marshall B., Edward Miguel, et al. (2009). "Warming Increases the Risk of Civil War in Africa". In: *Proceedings of the National Academy of Sciences* 106.49, pp. 20670–20674. DOI: 10.1073/pnas.0907998106.

Cerqueira Leite, Rogério Cezar de et al. (2009). "Can Brazil Replace 5% of the 2025 Gasoline World Demand with Ethanol?" In: *Energy* 34.5, pp. 655–661. DOI: 10.1016/j.energy.2008.11.001.

Chavez-Rodriguez, Mauro F. and Silvia A. Nebra (2010). "Assessing GHG Emissions, Ecological Footprint, and Water Linkage for Different Fuels". In: *Environmental Science & Technology* 44.24, pp. 9252–9257. DOI: 10.1021/es101187h.

Chow, Gregory C. (1957). *Demand for Automobiles in the United States: A Study in Consumer Durables*. North-Holland Pub. Co. 110 pp.

CME Group (2015). *Brent Crude Oil Futures Quotes*. URL: http://www.cmegroup.com/trading/energy/crude-oil/brent-crude-oil.html (visited on 01/06/2015).

Cohn, Avery S. et al. (2014). "Cattle Ranching Intensification in Brazil Can Reduce Global Greenhouse Gas Emissions by Sparing Land from Deforestation". In: *Proceedings of the National Academy of Sciences* 111.20, pp. 7236–7241. DOI: 10.1073/pnas.1307163111.

Cooper, Joseph, A. Nam Tran, and Steven Wallander (2017). "Testing for Specification Bias with a Flexible Fourier Transform Model for Crop Yields". In: *American Journal of Agricultural Economics* 99.3, pp. 800–817. DOI: 10.1093/ajae/aaw084.

Costinot, Arnaud, Dave Donaldson, and Cory B. Smith (2014). *Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7*

Million Fields around the World. Working Paper 20079. National Bureau of Economic Research. URL: http://www.nber.org/papers/w20079.

Cros, Marie-Josée et al. (2003). "A Biophysical Dairy Farm Model to Evaluate Rotational Grazing Management Strategies". In: *Agronomie* 23.2, pp. 105–122. DOI: 10.1051/agro:2002071.

De Vries, Bert J. M., Detlef P. van Vuuren, and Monique M. Hoogwijk (2007). "Renewable Energy Sources: Their Global Potential for the First-Half of the 21st Century at a Global Level: An Integrated Approach". In: *Energy Policy* 35.4, pp. 2590–2610. DOI: 10.1016/j.enpol.2006.09.002.

Deryugina, Tatyana and Solomon M. Hsiang (2014). *Does the Environment Still Matter? Daily Temperature and Income in the United States*. Working Paper 20750. National Bureau of Economic Research. DOI: 10.3386/w20750.

Deschênes, Olivier and Michael Greenstone (2007). "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather". In: *American Economic Review* 97.1, pp. 354–385. DOI: 10.1257/aer.97.1.354.

— (2011). "Climate Change, Mortality, and Adaptation: Evidence from Annual Fluctuations in Weather in the US". In: *American Economic Journal: Applied Economics* 3.4, pp. 152–185. DOI: 10.1257/app.3.4.152.

— (2012). "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Reply". In: *American Economic Review* 102.7, pp. 3761–73. DOI: 10.1257/aer.102.7.3761.

Dixit, Avinash K. and Robert S. Pindyck (1994). *Investment under Uncertainty*. Princeton University Press. 476 pp.

Doudna, Jennifer A. and Emmanuelle Charpentier (2014). "Genome Editing. The New Frontier of Genome Engineering with CRISPR-Cas9". In: *Science* 346.6213, p. 1258096. DOI: 10.1126/science.1258096.

Energy Information Administration (2014). *Annual Energy Outlook 2014*. Energy Information Administration. URL: http://www.eia.gov/forecasts/aeo/index.cfm.

ESALQ-LOG (2013). *Sifreca: Sistema de Informações de Fretes*. Yearbook. Piracicaba, São Paulo: ESALQ-LOG.

Fisher, Anthony C. et al. (2012). "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment". In: *American Economic Review* 102.7, pp. 3749–60. DOI: 10.1257/aer.102.7.3749.

Fisher, Irving (1907). *The Rate of Interest: Its Nature, Determination, and Relation to Economic Phenomena*. New York: The MacMillan Company. URL: https://mises.org/library/rate-interest-its-nature-determination-and-relation-economic-phenomena.

Food and Agriculture Organization of the United Nations (2014). *FAOSTAT*.

Garber, Steven and Steven Klepper (1980). "Extending the Classical Normal Errors-in-Variables Model". In: *Econometrica* 48.6, pp. 1541–1546. DOI: 10.2307/1912823.

Ghysels, Eric, Arthur Sinko, and Rossen Valkanov (2007). "MIDAS Regressions: Further Results and New Directions". In: *Econometric Reviews* 26.1, pp. 53–90. DOI: 10.1080/07474930600972467.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton (2013). "Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation". In: *Review of Environmental Economics and Policy* 7.1, pp. 23–46. DOI: 10.1093/reep/res015.

Griliches, Zvi (1957). "Specification Bias in Estimates of Production Functions". In: *Journal of Farm Economics* 39.1, pp. 8–20. DOI: 10.2307/1233881.

Griliches, Zvi and Jerry A. Hausman (1986). "Errors in Variables in Panel Data". In: *Journal of Econometrics* 31.1, pp. 93–118. DOI: 10.1016/0304-4076(86)90058-8.

Hanslow, Kevin et al. (2014). "Economic Impacts of Climate Change on the Australian Dairy Sector". In: *Australian Journal of Agricultural and Resource Economics* 58.1, pp. 60–77. DOI: 10.1111/1467-8489.12021.

Hausman, Jerry A. (2001). "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left". In: *The Journal of Economic Perspectives* 15.4, pp. 57–67.

Hijmans, Robert J. et al. (2005). "Very High Resolution Interpolated Climate Surfaces for Global Land Areas". In: *International Journal of Climatology* 25.15, pp. 1965–1978. DOI: 10.1002/joc.1276.

Holland, Stephen P., Jonathan E. Hughes, and Christopher R. Knittel (2009). "Greenhouse Gas Reductions under Low Carbon Fuel Standards?" In: *American Economic Journal: Economic Policy* 1.1, pp. 106–46. DOI: 10.1257/pol.1.1.106.

Holland, Stephen P., Jonathan E. Hughes, Christopher R. Knittel, and Nathan C. Parker (2014). "Some Inconvenient Truths about Climate Change Policy: The Distributional Impacts of Transportation Policies". In: *Review of Economics and Statistics* 97.5, pp. 1052–1069. DOI: 10.1162/REST_a_00452.

Hoogwijk, Monique Maria (2004). "On the Global and Regional Potential of Renewable Energy Sources". Ph.D Dissertation. Utrecht, The Netherlands: Utrecht University. 256 pp.

Houser, Trevor et al. (2015). *Economic Risks of Climate Change: An American Prospectus*. Columbia University Press.

Hsiang, Solomon M. (2016). "Climate Econometrics". In: *Annual Review of Resource Economics* 8.1, pp. 43–75. DOI: 10.1146/annurev-resource-100815-095343.

Hsiang, Solomon M. and Marshall B. Burke (2013). *Climate, Conflict, and Social Stability: What Does the Evidence Say?* SSRN Scholarly Paper ID 2302245. Rochester, NY: Social Science Research Network. URL: http://papers.ssrn.com/abstract=2302245.

Hurt, Chris, Wally Tyner, and Otto Doering (2006). *Economics of Ethanol*. ID-339. Lafayette: Purdue University.

IIASA and FAO (2012). *Global Agro-Ecological Zones (GAEZ v3.0)*. Laxenburg, Austria and Rome, Italy: IIASA/FAO.

International Energy Administration (2015). *Oil Market Report*. URL: https://www.iea.org/oilmarketreport/omrpublic/ (visited on 01/06/2015).

International Energy Agency (2015). *CO2 Emissions From Fuel Combustion*. Paris: International Energy Agency. URL: https://www.iea.org/publications/freepublications/publication/co2-emissions-from-fuel-combustion-highlights-2015.html.

Kellogg, Ryan (2014). "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling". In: *The American Economic Review* 104.6, pp. 1698–1734. DOI: 10.1257/aer.104.6.1698.

Khanna, Madhu, Basanta Dhungana, and John Clifton-Brown (2008). "Costs of Producing Miscanthus and Switchgrass for Bioenergy in Illinois". In: *Biomass and Bioenergy* 32.6, pp. 482–493. DOI: 10.1016/j.biombioe.2007.11.003.

Klepper, Steven and Edward E. Leamer (1984). "Consistent Sets of Estimates for Regressions with Errors in All Variables". In: *Econometrica* 52.1, pp. 163–183. DOI: 10.2307/1911466.

Levi, Maurice D. (1973). "Errors in the Variables Bias in the Presence of Correctly Measured Variables". In: *Econometrica* 41.5, pp. 985–986. DOI: 10.2307/1913819.

Lobell, David B. (2013). "Errors in Climate Datasets and Their Effects on Statistical Crop Models". In: *Agricultural and Forest Meteorology*. Agricultural prediction using climate model ensembles 170, pp. 58–66. DOI: 10.1016/j.agrformet.2012.05.013.

Lobell, David B. and Senthold Asseng (2017). "Comparing Estimates of Climate Change Impacts from Process-Based and Statistical Crop Models". In: *Environmental Research Letters* 12.1, p. 015001. DOI: 10.1088/1748-9326/aa518a.

McCallum, B. T. (1972). "Relative Asymptotic Bias from Errors of Omission and Measurement". In: *Econometrica* 40.4, pp. 757–758. DOI: 10.2307/1912970.

Mello, Francisco F. C. et al. (2014). "Payback Time for Soil Carbon and Sugar-Cane Ethanol". In: *Nature Climate Change* 4.7, pp. 605–609. DOI: 10.1038/nclimate2239.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw (1994). "The Impact of Global Warming on Agriculture: A Ricardian Analysis". In: *The American Economic Review* 84.4, pp. 753–771.

Mills, Edwin S. (1967). "An Aggregative Model of Resource Allocation in a Metropolitan Area". In: *The American Economic Review* 57.2, pp. 197–210.

Moraes, Márcia Azanha Ferraz Dias de and David Zilberman (2014). *Production of Ethanol from Sugarcane in Brazil*. Cham: Springer International Publishing. URL: http://link.springer.com/10.1007/978-3-319-03140-8.

Mukherjee, Deep, Boris E. Bravo-Ureta, and Albert De Vries (2013). "Dairy Productivity and Climatic Conditions: Econometric Evidence from South-Eastern United States". In: *Australian Journal of Agricultural and Resource Economics* 57.1, pp. 123–140. DOI: 10.1111/j.1467-8489.2012.00603.x.

Muth, Richard F. (1969). *Cities and Housing*. Chicago: University of Chicago Press.

Nelson, Gerald C. (2010). "Are Biofuels the Best Use of Sunlight?" In: *Handbook of Bioenergy Economics and Policy*. Ed. by Madhu Khanna, Jürgen Scheffran, and David Zilberman.

Natural Resource Management and Policy 33. Springer New York, pp. 15–25. URL: http://link.springer.com/chapter/10.1007/978-1-4419-0369-3_2.

Parker, Nathan C. (2011). "Modeling Future Biofuel Supply Chains Using Spatially Explicit Infrastructure Optimization". Ph.D. United States – California: University of California, Davis. 274 pp. URL: http://search.proquest.com/docview/872860188/abstract/B7DD7110F9A54DDEPQ/3?accountid=14496.

Porteous, A. S., R. E. Basher, and J. E. Salinger (1994). "Calibration and Performance of the Single-Layer Soil Water Balance Model for Pasture Sites". In: *New Zealand Journal of Agricultural Research* 37.1, pp. 107–118. DOI: 10.1080/00288233.1994.9513047.

Programa de Educação Continuada em Economia e Gestão de Empresas (2012). *Production Costs of Sugarcane, Sugar, and Ethanol in Brazil: 2011/2012 Season Report - South-Central [Portuguese: Custos de Produção de Cana-de-Açúcar, Açúcar e Etanol No Brasil: Acompanhamento Da Safra 2011/2012 - Centro-Sul]*. São Paulo: University of São Paulo.

Qi, L., B. E. Bravo-Ureta, and V. E. Cabrera (2015). "From Cold to Hot: Climatic Effects and Productivity in Wisconsin Dairy Farms". In: *Journal of Dairy Science*. DOI: 10.3168/jds.2015-9536.

Renewable Fuels Association (2014). *Annual Industry Outlook*. URL: http://www.ethanolrfa.org/pages/annual-industry-outlook (visited on 01/07/2015).

Rhoads, M. L. et al. (2009). "Effects of Heat Stress and Plane of Nutrition on Lactating Holstein Cows: I. Production, Metabolism, and Aspects of Circulating Somatotropin1". In: *Journal of Dairy Science* 92.5, pp. 1986–1997. DOI: 10.3168/jds.2008-1641.

Richards, Peter D., Robert T. Walker, and Eugenio Y. Arima (2014). "Spatially Complex Land Change: The Indirect Effect of Brazil's Agricultural Sector on Land Use in Amazonia". In: *Global Environmental Change* 29, pp. 1–9. DOI: 10.1016/j.gloenvcha.2014.06.011.

Rogelj, Joeri, Malte Meinshausen, and Reto Knutti (2012). "Global Warming under Old and New Scenarios Using IPCC Climate Sensitivity Range Estimates". In: *Nature Climate Change* 2.4, pp. 248–253. DOI: 10.1038/nclimate1385.

Schauberger, Bernhard et al. (2017). "Consistent Negative Response of US Crops to High Temperatures in Observations and Crop Models". In: *Nature Communications* 8, p. 13931. DOI: 10.1038/ncomms13931.

Schlenker, Wolfram, W. Michael Hanemann, and Anthony C. Fisher (2005). "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach". In: *The American Economic Review* 95.1, pp. 395–406.

Schlenker, Wolfram and Michael J. Roberts (2009). "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change". In: *Proceedings of the National Academy of Sciences* 106.37, pp. 15594–15598. DOI: 10.1073/pnas.0906865106.

Searchinger, Tim et al. (2013). *Creating a Sustainable Food Future: Interim Findings — World Resources Institute*. Washington, DC: World Resources Institute. URL: http://www.wri.org/publication/creating-sustainable-food-future-interim-findings.

Slade, Raphael, Ausilio Bauen, and Robert Gross (2014). "Global Bioenergy Resources". In: *Nature Climate Change* 4.2, pp. 99–105. DOI: 10.1038/nclimate2097.

Somerville, Chris et al. (2010). "Feedstocks for Lignocellulosic Biofuels". In: *Science* 329.5993, pp. 790–792. DOI: 10.1126/science.1189268.

Soybean And Corn Advisor (2014). *Irrigation Slowly Expanding in Brazil, 5% of Grain Crop Irrigated*. URL: http://www.soybeansandcorn.com/news/Jan14_12-Irrigation-Slowly-Expanding-in-Brazil-5-of-Grain-Crop-Irrigated (visited on 02/03/2015).

State of California (2009). *Low Carbon Fuel Standard*.

Theil, H. (1957). "Specification Errors and the Estimation of Economic Relationships". In: *Revue de l'Institut International de Statistique / Review of the International Statistical Institute* 25 (1/3), pp. 41–51. DOI: 10.2307/1401673.

US EPA Climate Change Division (2013). *Social Cost of Carbon*. EPA and other federal agencies use the social cost of carbon (SCC) to estimate the climate benefits of rulemakings. URL: http://www.epa.gov/climatechange/EPAactivities/economics/scc.html (visited on 03/17/2014).

Waclawovsky, Alessandro J. et al. (2010). "Sugarcane for Bioenergy Production: An Assessment of Yield and Regulation of Sucrose Content". In: *Plant Biotechnology Journal* 8.3, pp. 263–276. DOI: 10.1111/j.1467-7652.2009.00491.x.

Welch, Jarrod R. et al. (2010). "Rice Yields in Tropical/Subtropical Asia Exhibit Large but Opposing Sensitivities to Minimum and Maximum Temperatures". In: *Proceedings of the National Academy of Sciences* 107.33, pp. 14562–14567. DOI: 10.1073/pnas.1001222107.

West, J. W. (2003). "Effects of Heat-Stress on Production in Dairy Cattle". In: *Journal of Dairy Science* 86.6, pp. 2131–2144. DOI: 10.3168/jds.S0022-0302(03)73803-X.

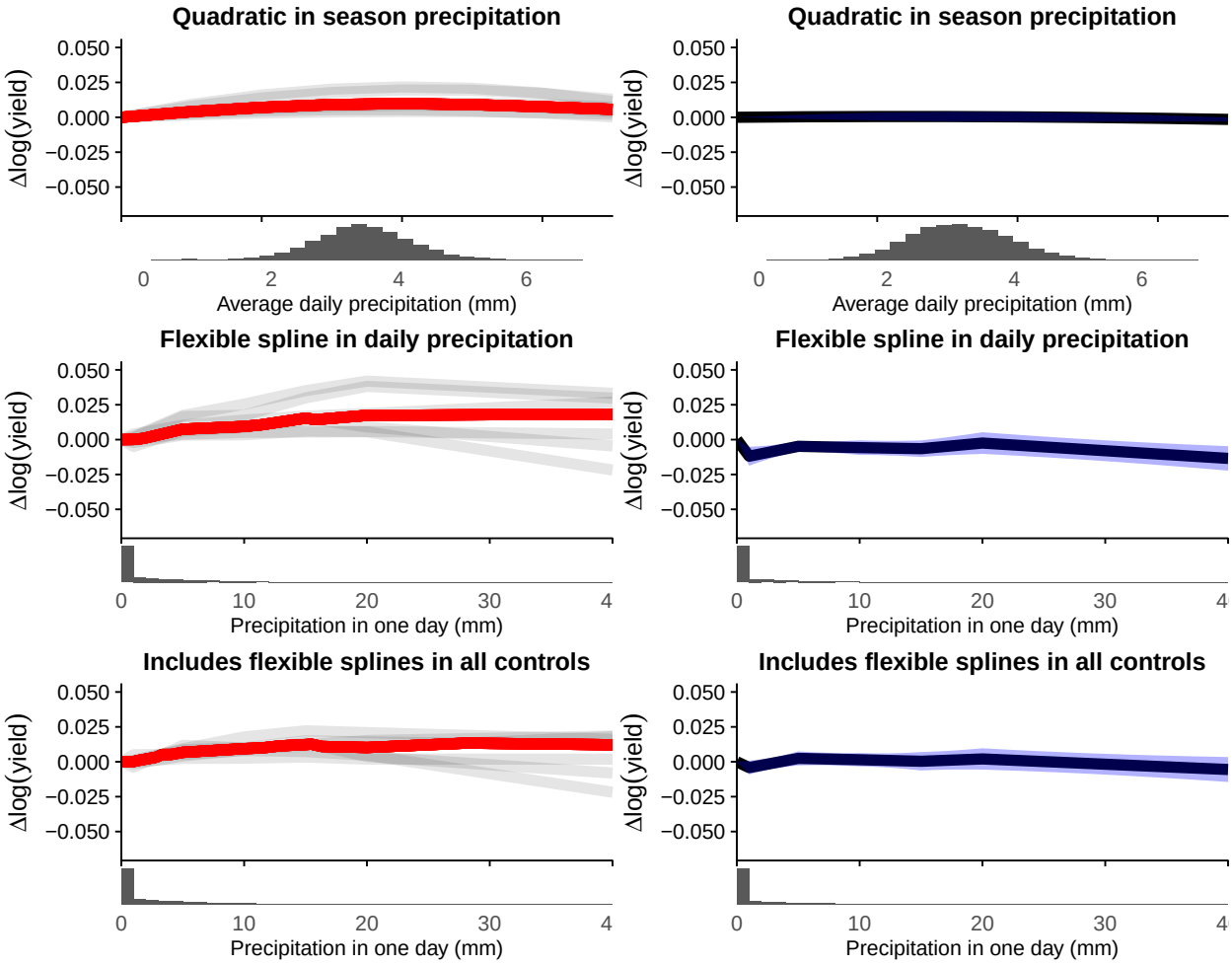
Wickens, Michael R. (1972). "A Note on the Use of Proxy Variables". In: *Econometrica* 40.4, pp. 759–761. DOI: 10.2307/1912971.

Wirsenius, Stefan (2003). "The Biomass Metabolism of the Food System: A Model-Based Survey of the Global and Regional Turnover of Food Biomass". In: *Journal of Industrial Ecology* 7.1, pp. 47–80. DOI: 10.1162/108819803766729195.

Appendix A

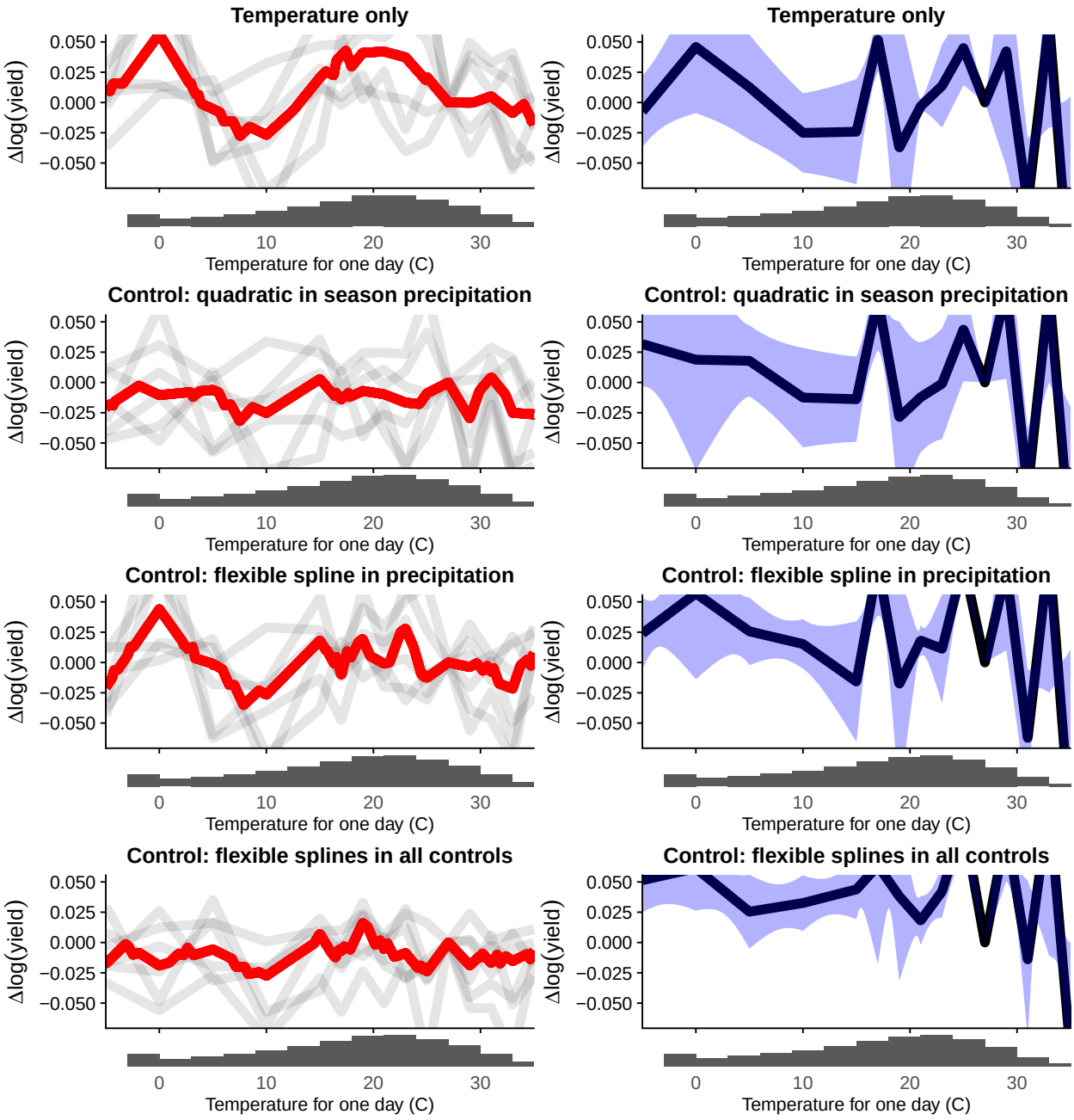
Appendix to "Does the 'Iron Law' always hold? The impact of measurement error in climate econometrics"

Figure A1: Process model and empirical model precipitation response functions using flexible temperature



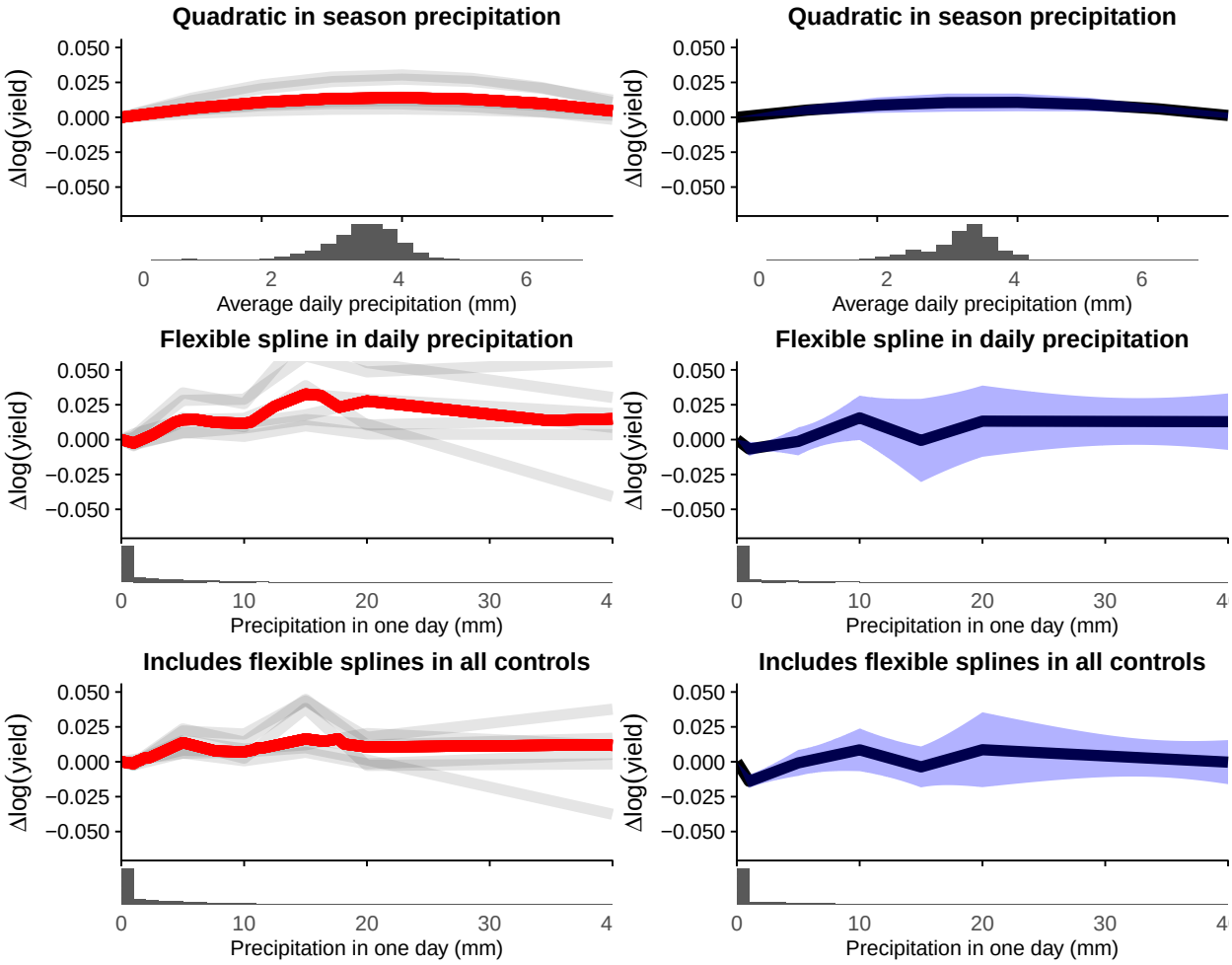
This figure plots Equations (2.10) *and* (2.11) *for a single day of precipitation, using Equation* (2.16) *for temperature. Each plot is vertically centered so that change in log(yield) takes a value of zero when precipitation is 0mm. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of precipitation. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of daily precipitation.*

Figure A2: Process model and empirical model flexible temperature response functions in the long differences model



This figure plots Equation (2.16) *for a single day of temperature. Each plot is vertically centered so that change in log(yield) takes a value of zero when temperature is 27° C. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of temperature. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of the full time series of temperature using single sine interpolation.*

Figure A3: Process model and empirical model precipitation response functions using flexible temperature in the long differences model



This figure plots Equations (2.10) *and* (2.11) *for a single day of precipitation, using Equation* (2.16) *for temperature. Each plot is vertically centered so that change in log(yield) takes a value of zero when precipitation is 0mm. The first column plots the results for the statistically emulated process crop models; the light gray lines represent each individual model and the red line is the median of the plotted points for each value of precipitation. The second column plots the results for the empirical model with 95% confidence bands calculated assuming error clustering by state and year. Subplot titles describe which control variables are included. Below each plot is a histogram of daily precipitation.*

Appendix B

Appendix to "Empirical estimation of the impact of weather on dairy production"

B.1 Calculation of the contribution of pasture to global caloric production

Table 3.3 ranks land uses by their contribution to global caloric production; this section fully describes the calculations for this table. The FAO food balance data includes:

1. "Food supply": the quantity of food available for human consumption net of food-system waste, feed utilization, and changes in storage.
2. "Feed": the quantity of the food product utilized as animal feed.
3. "Production": the quantity of new production of the food product.

The data reports each of these items by "food type"; food types include items such as "Wheat and products", and "Poultry Meat". All values are for the most recent year, 2011. The FAO food balance data allows me to calculate the contribution of each land use to the food supply via plant-based foods directly, as these are simply the "food supply" values. Assuming a feed conversion ratio for each feed product also allows me to calculate the contribution of each plant-based land use *that is measured in the food balance data* to the food supply via animals. However, the contributions from both pasture and crop residues are not directly measured in the food balance data, as these are not also food products.

I am, however, able to measure the total contribution to the food supply for each animal product. With an assumed feed conversion ratio, along with the feed quantities, I can calculate the total quantity of animal calories that can be attributed to those feed products. I then calculate the contribution of pasture and crop residues as the residual of this relationship.

Table A1: Typical feed conversion ratios for land animal food products

Food Item	Feed conversion ratio
Poultry Meat	0.11
Pigmeat	0.10
Butter, Ghee	0.07
Milk - Excluding Butter	0.07
Cream	0.07
Bovine Meat	0.01
Eggs	0.13
Mutton & Goat Meat	0.01

Explicitly, my calculation is as follows. Suppose Q_i^F is the "food supply" of food type i, Q_i^A is the feed utilized of food type i, both measured in kcal/capita/day. Total "plated" food supply, both directly through plant-based products, and indirectly through animals, for food type i, is then:

$$Q_i^H = Q_i^F + \alpha_i Q_i^A \tag{B.1}$$

where i indexes plant-based food/feed types (e.g. Maize and Products), α_i is the average feed conversion ratio for food/feed type i. I calculate the total contribution of plant food products that are measured in the FAO food balance data to animal calories as:

$$\sum_{i \in I} \alpha_i Q_i^A, \tag{B.2}$$

for $I = \{$Apples and products; Bananas; Barley and products; Beans; Bovine Meat; Butter, Ghee; Cassava and products; Cereals, Other; Cocoa Beans and products; Coconuts - Incl Copra; Dates; Eggs; Fruits, Other; Groundnuts (Shelled Eq); Maize and products; Meat, Other; Milk - Excluding Butter; Millet and products; Mutton & Goat Meat; Oats; Offals; Offals, Edible; Oilcrops Oil, Other; Oilcrops, Other; Onions; Oranges, Mandarines; Peas; Plantains; Potatoes and products; Pulses, Other and products; Rape and Mustard Oil; Rape and Mustardseed; Rice (Milled Equivalent); Roots, Other; Rye and products; Sesame seed; Sorghum and products; Soyabean Oil; Soyabeans; Stimulants; Sugar (Raw Equivalent); Sugar cane; Sugar non-centrifugal; Sunflower seed; Sweet potatoes; Sweeteners, Other; Tomatoes and products; Vegetables, Other; Wheat and products; Yams$\}$. Note that this excludes fish feed types, as these are primarily used to produce other fish food products.

As I can not determine which animals are fed which feeds, I assume $\alpha_i = \alpha \quad \forall i$. To calculate the average feed conversion ratio, I collect typical feed conversion ratios by animal from Searchinger et al. (2013, p. 37), which I show in Table A1. Next, I calculate the animal feed consumed for each land-based animal food product j using:

$$Q^{A,j} = Q_j^P / \alpha^j \tag{B.3}$$

where α^j is the feed conversion ratio for animal product j and $Q^{P,j}$ is the production of animal product j. $Q^{A,j}$ then represents the total quantity of feed across all feed types, for animal product j. Finally, to obtain α, the average conversion ratio, I use:

$$\alpha = \frac{\sum_{j \in J} Q^{P,j}}{\sum_{j \in J} Q^{A,j}} \tag{B.4}$$

To obtain the contribution from pasture and crop residues, I use the identity:

$$\sum_{j \in J} Q^{A,j} = \sum_{i \in I} (Q_i^A) + Q_{Pasture}^A \tag{B.5}$$

which says that the total amount of feed consumed by land animals is equal to the total amount of feed consumed from plant food products, $\sum_{i \in I}(Q_i^A)$, plus the total amount of feed consumed as pasture and crop residues, $Q_{Pasture}^A$. Finally, I calculate the total contribution to the food supply from pasture and crop residues as $Q_{Pasture}^H = Q_{Pasture}^A * \alpha$. I then calculate the final proportions by dividing each Q_i^H by the "Grand Total" food supply value in the FAO data.

The relative contribution of pasture and crop residues is approximately 70% and 30% respectively (Wirsenius 2003).

B.2 Climate change projections

In order to obtain indicative changes in production under climate change, I follow Houser et al. (2015) by simulating future daily weather by randomly sampling historical weather and adding differences generated by a climate model. In the current version of this project, I use a single climate model, HadGEM2-ES.

The following method for computing downscaled future climate projections differs from that in Houser et al. (2015) Secondly, instead of computing weather changes over fixed finite periods (e.g. 1981-2000 to 2040-2059), I compute weather changes and sum the projected impacts for all years.

Other than using a single climate model, my method differs from that in Houser et al. (2015) in two ways. Firstly, I compute projections for all years out to 2100. Secondly, I filter the climate model output using a LOWESS smoother, so that only the first-order trends and decadal variation from the climate model are used.

The full process is as follows: For climate model/emissions scenario/realization i:

1. Extract monthly data for i for all grid cells that overlap with the weather grid cells used in the analysis, from 30 years before the start of the analysis to 2100. Weather grid cells that do not overlap any cell in i are matched to the nearest cell in i. Index the grid cells in i by j.

2. For i, month-of-year, variable, and grid cell j, compute lowess-smoothed monthly data with a smoother span of 0.3[1]. This generates a smoothed path of each variable that preserves decadal variation.

3. For each future year t available in i, randomly select a year s from the weather data available (for NZ, this is 1972-2015). Then for every weather grid cell k in climate model grid cell j, and month in t, add the monthly difference from s to t using the lowess smoothed data.[2]

Soil moisture projections

In CMIP5, soil moisture variables available include total soil moisture across all layers in kg/m^2 and soil moisture in the top 0.1m in kg/m^2. However, neither of these soil moisture units exactly match those in the VCSN. In addition, some climate models do not reproduce the timing of seasonality of soil moisture as in reality. Thus, to obtain the mapping between the soil moisture units in the VCSN and the climate models, I make the assumption that the magnitude, but not necessarily the timing of the seasonality in soil moisture in the climate models is correct. This exploits the largest source of variation in soil moisture that is common between the weather data and the climate model data, the seasons. To operationalize this, I employ the following relationship in constructing the future soil moisture data:

$$\widehat{SM}_{it}^{W} = \min(\overline{SM}^{W}) + \frac{\max(\overline{SM}^{W}) - \min(\overline{SM}^{W})}{\max(\overline{SM}_{i}^{C}) - \min(\overline{SM}_{i}^{C})} * \left(SM_{it}^{C} - \min(\overline{SM}_{i}^{C})\right) \tag{B.6}$$

where i indexes climate models, t indexes future days, W indicates the weather data, C indicates the climate model data, and averages are computed over all historical years which exist in both the weather data and the climate model data.

[1] In `R`, this is computed as `lowess(x = month, y = value, f = 0.3)$y`.

[2] In the New Zealand data, years are defined to run from June 1 to May 31. 366-day future years which are matched with 365 day past years use day 365 twice.

Appendix C

Appendix to "The potential for renewable fuels under greenhouse gas pricing: The case of sugarcane"

C.1 Parameter values

Table A1 presents all parameter inputs into the refinery-level NPV calculation. "*" denotes variables that, within reasonable bounds, materially affect the final results.

Table A1: Sugarcane investment model parameter values

Parameter	Value	Source
Refinery capacity (m^3/year)	250,000	Assumed.
Construction costs per refinery ($R million)	680.9	Bloomberg New Energy Finance. Predicted value for 2014 from a regression of construction cost on capacity and a linear trend using 22 ethanol-only refineries built in Brazil from 2005-2014.
Real discount rate (%)	6.1%*	Petrobras' WACC from wikiwealth.com, less expected inflation from tradingeconomics.com
Sugarcane cycle length (years)	6	Assumed.
Lifespan of each refinery (years)	30*	Assumed.
Additional maintenance costs as a proportion of initial construction costs (%)	0*	Assumed. Note that maintenance costs are explicitly accounted for in the refinery operating costs.
Construction period (year)	1	Assumed.

Ethanol price	R$1840/m3*	Nearest upcoming hydrous ethanol futures price on BM&F Bovespa (Collected October 25, 2016) (Brazilian Securities, Commodities and Futures Exchange 2014)).
Oil price	US$48.8-US$53.3 (2015-2088)	Nymex futures prices, averaged for each year. Unobserved years take the latest value. (Collected October 25, 2016 (CME Group 2015)).
Refinery operating costs	R$278/m3	PECEGE expansion region costs of refinery production less capital costs, depreciation, and rent (2012)
Feedstock costs	R$3189/ha	PECEGE expansion region costs of cane production less capital costs, depreciation, rent, and transportation costs (2012).
Feedstock transport costs	R$8.2/Mg Cane	Sugarcane transportation costs from Bonomi et al. (2012).
Pasture to sugarcane conversion cost	R$181/ha	Conversion cost used in IBGE (2013).
Annual yield increases	0.32Mg/ha	Projected from historical trend using Brazilian Sugarcane Industry Association (2014).
Ethanol yield per Mg sugarcane stalk	$0.086 m^3$/Mg Cane	Calculated from Somerville et al. (2010).

Table A2 presents all parameter inputs into refinery and pasture allocation optimization model.

Table A2: Optimization model constraint values

Parameter	Value	Source
Maximum number of refineries built per year	36	Assumed to be equivalent to 50% more than on a production capacity basis.
First year in the model	2014	Assumed.
Final potential build start year	2038	Assumed.
Proportion of total pasture available for conversion.	0.5	Assumed.

Table A3: GDP deflators for USD and BRL

USD		BRL	
2000	89.02	2000	100.00
2001	91.05	2001	108.09
2002	92.45	2002	118.80
2003	94.29	2003	135.38
2004	96.88	2004	145.91
2005	100.00	2005	156.84
2006	103.07	2006	167.34
2007	105.82	2007	178.05
2008	107.89	2008	193.83
2009	108.71	2009	208.09
2010	110.04	2010	225.91
2011	112.31	2011	244.69
2012	114.33	2012	259.04
2013	116.03	2013	275.90
2014	117.72	2014	294.93

Note: Collected from the World Bank Development Indicators.

Table A4: Miscellaneous parameter values

Parameter	Value	Source
Soil carbon change when converting from pasture to sugarcane (Mg/ha)	31.8	Mello et al. (2014)

C.2 Oil and ethanol market equilibrium

In this subsection, we describe the simple energy market equilibrium model we use to generate the effect of a global GHG price, or equivalent policy, on the producer prices for oil and ethanol. Because we use these equations exactly, we also provide our particular parametrization.

The key assumptions we employ are: oil supply, ethanol supply, and energy demand are constant elasticity functions of price, oil and ethanol are perfect substitutes in the energy market, BTUs are the only valued component of either fuel, and oil BTUs are penalized using a scalar multiplier to equate the initial prices of the two fuels on a per-BTU basis.

The equilibrium equations are given below. We use subscript t's to denote variables that

potentially change over time in our later simulation.

$$P^s_{ot} = P^d_{BTU,t} BTU^*_o - P_{GHG,t} GHG_o \tag{C.1}$$
$$P^s_{et} = P^d_{BTU,t} BTU_e - P_{GHG,t} GHG_e \tag{C.2}$$
$$Q_{BTU,t} = A_d (P^d_{BTU,t})^{-r} \tag{C.3}$$
$$Q_{BTU,t} = Q_{ot} BTU^*_o + Q_{et} BTU_e \tag{C.4}$$
$$Q_{ot} = A_o (P^s_{ot})^{\eta_o} \tag{C.5}$$
$$Q_{et} = A_e (P^s_{et})^{\eta_e} \tag{C.6}$$

In the above, P^s_{ot} is the price per barrel of oil to suppliers, P^s_{et} is the price of ethanol per m^3 to suppliers, $P^d_{BTU,t}$ is the price of world energy demanded in ethanol-equivalent quadrillion BTUs,[1] BTU^*_o is the energy content of oil per barrel, scaled to equate the initial prices of the fuels on a per-BTU basis, BTU_e is the energy content of ethanol per m^3, $P_{GHG,t}$ is the price of GHGs emitted, GHG_o is the GHGs emitted per-unit oil, GHG_e is the GHGs emitted per unit ethanol, $Q_{BTU,t}$ is the world supply of ethanol-equivalent BTUs supplied across both fuels, $P^d_{BTU,t}$ is the price of an ethanol-equivalent BTU, Q_{ot} is the world supply of oil in barrels, Q_{et} is the world supply of ethanol in m^3. Prices and quantities are the endogenous variables in this system; the calculation of the remaining parameters is summarized in table A5:

Table A5: Oil and ethanol market parameter calculations

Parameter	Formula	Value	Explanation
$P_{GHG,t}$		\$38.94–\$96.79 (2014–2068)	US federal social cost of carbon, linearly interpolated/extrapolated (US EPA Climate Change Division 2013).
GHG_o		0.657	Lifecycle GHG content of a barrel of crude oil (Chavez-Rodriguez and Nebra 2010).
GHG_e		0.256	Lifecycle GHG content of an m^3 of ethanol.[2]

[1] The units in this model are chosen both for computational and expositional purposes.

[2] http://www.arb.ca.gov/fuels/lcfs/CleanFinalRegOrder112612.pdf

A_d	$\frac{Q_e^0 BTU_e + Q_o^0 BTU_o^*}{(P_e^0/BTU_e)^{-r}}$	$\frac{2.24 + 203 * BTU_o^*}{(26.2)^{-0.072}}$	Equation (C.3) rearranged using 2014 values. World oil supply in 2013Q4-2014Q3 (International Energy Administration 2015), Brent oil futures price in 2014 dollars for year t (CME Group 2015) (unavailable years are taken to be the final price in this list), r is the central value from (Holland, Hughes, and Knittel 2009).
A_{ot}	$\frac{Q_o^0}{(P_o^t)^{\eta_o}}$	$\frac{33.9}{(P_o^t)^{0.5}}$; $P_o^t \in$ {\$48.8–\$53.3}	World oil supply in 2013Q4-2014Q3 (International Energy Administration 2015), Brent oil futures price in 2014 dollars for year t (CME Group 2015) (unavailable years are taken to be the final price in this list), η_o is the central value from (Holland, Hughes, and Knittel 2009).
A_e	$\frac{Q_e^0}{(P_e^0)^{\eta_e}}$	$\frac{99.9}{(1840)^1}$	World ethanol supply in 2013 (Renewable Fuels Association 2014), Ethanol futures for February 2015 delivery to SP (BM&F Bovespa 2015), η_e is the central value from (Holland, Hughes, and Knittel 2009).
BTU_{ot}^*	$BTU_o \frac{P_{ot}^0/BTU_o}{P_e^0/BTU_e}$	$\frac{49.8}{1840/22391726}$	Scales the BTU content of a barrel of oil in the model to equate the initial prices of the (scaled) oil BTUs and ethanol BTUs. P_o^0 (Energy Information Administration 2014; CME Group 2015; Brazilian Securities, Commodities and Futures Exchange 2014).

BTU_e	$\frac{BTU_e}{boe_e} * \text{bbl/m}^3 * 10^6$	$3.560 * 6.29 * 10^6$	BTUs per m^3 ethanol. Standard measure from (Energy Information Administration 2014).

The results using GHG prices from US EPA Climate Change Division (2013), oil prices from CME Group (2015), and a recent ethanol price are presented in Figure A1.

Figure A1: Greenhouse gas and ethanol prices produced using energy market equilibrium model