Climate Change and U.S. Agriculture: Accounting for Multi-dimensional Slope Heterogeneity in Production Functions

By Michael Keane and Timothy Neal*

Draft: January 2, 2020

We study potential impacts of future climate change on U.S. agricultural productivity using county-level yield and weather data from 1950 to 2015. To account for adaptation of production to different weather conditions, it is crucial to allow for both spacial and temporal variation in the production process mapping weather to crop yields. We present a new panel data estimation technique, called mean observation OLS (MO-OLS) that allows for spatial and temporal heterogeneity in all regression parameters (intercepts and slopes). Both forms of heterogeneity are important: We find strong evidence that production function parameters adapt to local climate, and also that sensitivity of yield to high temperature declined from 1950-89. We use our estimates to project corn yields to 2100 using 19 climate models and three greenhouse gas emission scenarios. We predict unmitigated climate change will greatly reduce yield. Our mean prediction (over climate models) is that adaptation alone can mitigate 36% of the damage, while emissions reductions consistent with the Paris targets would mitigate 76%.

JEL: C23, C54, D24, Q15, Q51, Q54, Q55

Leading scientific and environmental institutions warn that future climate change may severely impact agricultural productivity and global food supply (Porter et al. 2014). But studies that attempt to estimate sensitivity of agriculture to climate, in hopes of obtaining insight into effects of future climate change, have given mixed results. Projected impacts of climate change on U.S. agriculture in particular range from severe damage to productivity (e.g. Schlenker and Roberts 2009) to negligible damage (e.g. Butler and Huybers 2013).

* Keane: University of New South Wales, Sydney, NSW 2000 (email: m.keane@unsw.edu.au); Neal: Department of Economics, University of New South Wales, Sydney, NSW 2000 (email: timothy.neal@unsw.edu.au).

Resolving this uncertainty is one of the top priorities for improving climate change impact assessments (Lobell and Burke 2008).

The impact of climate change on crop yield will depend critically on adaptation by agricultural producers. This may entail using more heat tolerant hybrids, improved water retention in fields, irrigation, adjusting sowing density, etc. The existence of adaptation in the historical data implies spacial and temporal heterogeneity in the production process mapping weather to crop yields. Thus, we would expect the production function to exhibit region and time fixed effects in both intercepts and slopes, where the fixed effects are correlated with regional and temporal variation in climate. This complex heterogeneity structure creates important challenges for the proper econometric modelling of climate impacts - challenges that have not been fully addressed in prior literature.

We investigate these issues using weather and crop yield data for U.S. counties from 1950 to 2015. We focus on corn, as it is historically the largest crop in the U.S. in terms of tonnage. Prior to our work, several authors investigated agricultural adaptation using panel data regressions of yield on temperature and precipitation (see, e.g., Schlenker and Roberts 2009, Burke and Emerick 2016). Particularly relevant are Butler and Huybers (2013), who allow for heterogeneity in heat sensitivity across counties, and Roberts and Schlenker (2012), who allow for variation over time (common to all counties). But prior work has not accommodated adaptation across both regions and time.

Our first main contribution is to present a new method to estimate the extent of historical adaptation to high temperatures. Since variation in heat sensitivity can occur across space *and* over time, we develop a method we call 'mean observation OLS' (MO-OLS) that is able to feasibly estimate in large panel datasets a model that contains both county and time fixed effects in both the intercept and the slope coefficients. It is a more flexible approach for modelling adaptation than in prior studies, and we find that it offers substantial improvements in fit over existing econometric models of yield.[1]

For comparison, we also consider a simple parametric approach where, in the crop yield regression, we let the coefficient on high temperatures be a theoretically motivated non-linear function of temperature itself – allowing the sensitivity of crop yield to high

[1] In a follow up article, Keane and Neal (2018) show that a MO-OLS model with county/time heterogeneity in intercepts and slopes can forecast corn yield out-of-sample as accurately as a deep neural net using the same inputs.

temperatures to decline as the frequency of high temperatures increases. We find a close agreement between the parametric estimates of temperature sensitivity and the MO-OLS (fixed-effects) estimates. Both imply that heat sensitivity declines with temperature according to a log-linear relationship similar to that implied by our simple theory.[2]

Using both MO-OLS and the parametric approach, we find that significant adaptation occurred across hot/cool counties, as well as over time from 1950 to 1989.[3] But we find that adaptation stalled after 1989. This is consistent with a trend towards higher sowing densities, which generate higher yield in good years, but worsen heat stress in drought years (see Lobell et al. (2014)). The substantial expansion of crop insurance in the early 1990s may have encouraged this trend, by reducing the incentives for farmers to adapt to extreme heat (see Annan and Schlenker 2015).

Our second main contribution is to use our econometric models to project future corn yields annually until 2100, using temperature and precipitation projections from 19 climate models under three greenhouse gas emissions scenarios. We provide projections both with and without adaptation, utilizing our estimated relationship between temperature and heat sensitivity to predict adaptation. Our econometric models allow us to compare the effectiveness of adaptation vs. emissions reduction as ways to mitigate crop damage. By using 19 climate models we quantify the variability in outcomes across models.

The UN-IPCC uses a set of "representative concentration pathways" (RCPs) for greenhouse gas emissions based on different policy scenarios. We predict "business as usual" emissions growth (the RCP85 scenario), combined with no adaptation to climate change, will cause catastrophic damage to corn yield. Our central projection is a 70% reduction in yield by 2100, with an 80% prediction interval from 51 to 89%.[4] We predict that adaptation consistent with that observed in the historical data may avert from 29 to 44% of the total damage over 2020-2100. Thus, while adaptation may be important, the U.S. cannot rely on adaptation alone to protect corn yields from substantial impacts.

[2]Our approach is related to Butler and Huybers (2013), who let the sensitivity of yield to high temperature depend on average temperature in a county. But they only allow for heterogeneity in heat sensitivity across counties and not over time. We show that time effects are also very important.

[3]We also show, in Appendix B, that ignoring adaptation (i.e., ignoring fixed effects in slopes) in econometric models of crop yield leads to underestimation of yield sensitivity to high temperatures. Our estimates imply this underestimation is on the order of 60 to 85 percent.

[4]The wide prediction interval stems from substantial disagreement across climate models about future weather conditions in the corn growing counties of the US. In the results section we'll show that uncertainty about future yields is due more to uncertainty across climate models than across econometric models.

Next, we consider "moderate" emissions reductions, consistent with the UN-IPCC's "RCP45" scenario, which is somewhat more ambitious than current government pledges. Our central projection is that this would avert 55% of the total damage to corn yield without any adaptation, or 61% if we factor in adaptation. Finally, we consider substantial emissions reductions, consistent with the RCP26 scenario, which follows from the most ambitious targets under the Paris agreement. We predict this would avert 76% of damage, even without adaptation. Thus, while adaption has the potential to avert a meaningful fraction of yield damage, substantial emissions reductions are necessary to avert most of the potential damage from climate change.

The paper is structured as follows: Section I presents a simple model of agricultural yield with adaptation, to provide a coherent framework for the empirical work. Section II discusses our econometric methods, including the MO-OLS estimator. Section III describes our data. Section IV presents our main econometric results for corn yield. Section V presents our projections for corn yield through to 2100. Section VI concludes. A Mathematical Appendix provides proofs of MO-OLS properties. Eight online appendices extend the results of the article, including Monte Carlos and results for soybeans.

I. A Simple Model of Agricultural Yield with Weather and Adaptation

There is a long tradition in agricultural economics of estimating production functions for corn yield. Starting from the classic work by Wallace (1920), researchers have estimated regressions for annual yield as a function of temperature and precipitation during the growing season. Recent work uses modern panel-data techniques to control for county fixed effects (e.g., to account for soil quality) and common time effects. Here, we present a simple model that (i) rationalizes the conventional econometric specification and (ii) shows how adaptation generates fixed-effects heterogeneity in slopes.

We start with a production function for corn that incorporates measures of temperature:

$$Y_{it}/C_{it} = A_t \mu_i I_{it} \left(1 + \beta_1 (GDD_{it} - GDD_{min}) + \beta_2 KDD_{it}\right) \tag{1}$$

Here Y_{it} is output of corn for farmer i at time t and C_{it} is the number of acres planted, so the dependent variable is yield. A_t is total factor productivity at time t and μ_i is an area effect (e.g., soil quality). The term I_{it} is a composite of conventional inputs; e.g., capital, labor, fertilizer. We assume the functional form of I_{it} is common for all i.

The variables GDD_{it} and KDD_{it} capture effects of temperature. 'Growing degree days,' or GDD_{it}, is the total hours in the growing season that the crop experiences beneficial temperature. 'Killing degree days,' or KDD_{it}, is the total hours of harmful temperatures. GDD_{min} is the minimum level of GDD needed to obtain a positive yield. The percent shift in yield due to temperature factors is $x = \beta_1(GDD_{it} - GDD_{min}) + \beta_2 KDD_{it}$. We omit precipitation for simplicity, but it is included in the econometric models.

Now we show how the simple model in (1) can rationalize the yield models estimated in the literature. Taking the log of crop yield and using the approximation $ln(1+x) \approx x$, which is accurate, as values of x outside the +20% to -20% range are rare, we obtain:

$$y_{it} = ln(A_t) + (ln(\mu_i) - \beta_1 GDD_{min}) + ln(I_{it}) + \beta_1 GDD_{it} + \beta_2 KDD_{it} \tag{2}$$

where $y_{it} = ln(Y_{it}/C_{it})$. Equation (2) is akin to that estimated in several recent papers. Typically, these papers use fixed effects over i and t to capture the A_t, μ_i and I_{it} terms.

This approach is valid if we can write $ln(I_{it}) = f_i + f_t + \epsilon_{it}$, where f_i and f_t denote variation in inputs over farms/time that are captured by unit and time fixed effects,[5] while ϵ_{it} is an idiosyncratic factor uncorrelated with time t weather shocks. Then, ϵ_{it} provides the econometric error for estimation of (2). As a practical matter, research on corn yield emphasizes the role of weather in the production function (using i and t dummies to capture other inputs) for two reasons: variation in yield over time - beyond what is explained by farm/time effects - is well explained by variation in weather (Tannura et al. (2008), Westcott and Jewison 2013, Wang et al. 2012), and modification of conventional inputs after weather shocks are revealed has only minor effects on yield.[6]

Next, we extend this simple model to account for adaptation. Assume that by bearing a cost farmers can reduce sensitivity of yield to high temperature (e.g., paying a premium

[5] For instance, if farmers use a common technology and face common input and output prices, then the time factor f_t will capture proportional year-to-year shifts in inputs across all farmers in response to changes in input or expected output prices.

[6] In contrast, if changes in I_{it} could mitigate unanticipated shocks to KDD, then β_2 would be biased towards zero. The traditional agricultural economics literature (implicitly) rules this out when it estimates versions of (2). But (2) could still be interpreted as a reduced form, provided input prices are captured by the fixed effects.

for drought resistant seed). To capture this, let the KDD coefficient be $\beta_{2it} = s/(1+\alpha_{it})$, where $s < 0$ is the effect of high temperatures on yield absent any adaptation, while α_{it} denotes units of adaptation purchased by farmer i in period t. Letting γ denote the price of adaptation, profit for farmer i at time t is $\pi_{it} = p_t Y_{it} - \gamma\alpha_{it} - r_t I_{it}$ where p_t is the price of the crop, and r_t is the rental rate per unit of I_{it}.[7] To maximize profit, farmers purchase the optimal level of adaptation. Setting $\partial\pi/\partial\alpha = 0$ we obtain:

$$\alpha^*_{it} = \sqrt{\frac{p_t(C_{it}A_t\mu_i I_{it})(-s)KDD_{it}}{\gamma}} - 1 \tag{3}$$

Thus as KDD_{it} increases the optimal level of adaptation increases. Farms in hotter counties/time periods have more incentive to adapt. Figure 1 plots the optimal level of adaptation α^*_{it} and the implied coefficient β_{2it} against KDD_{it}. As we see, the relationship between KDD_{it} and the heterogeneous coefficient β_{2it} closely resembles a log-linear function. Of course, this depends on the functional form in (2). However, this prediction of the simple model is testable, and it is supported by our estimates. More importantly, the model illustrates how adaptation renders the coefficient on KDD_{it} heterogeneous, and the heterogeneity will be correlated with the regressor itself. To capture this requires an econometric method that allows for fixed effects in slopes (not just intercepts).

Letting *both* the KDD_{it} and GDD_{it} coefficients be heterogeneous across i and t, and using time and area fixed-effects (which we denote by c_t and c_i) to pick up the A_t, μ_i, f_t and f_i terms, we obtain a modified version of equation (2) of the form:

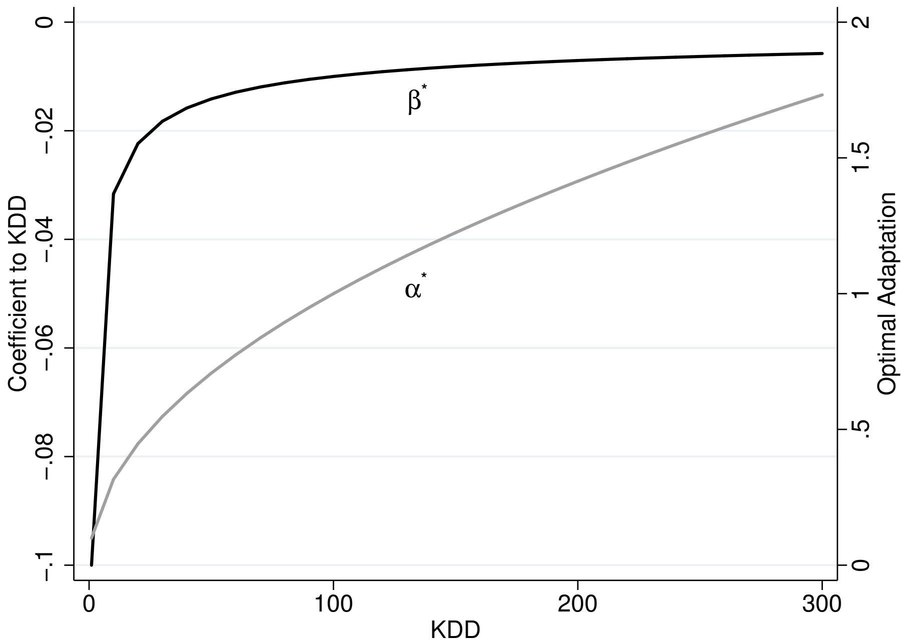
$$y_{it} = c_t + c_i + \beta_{1it}GDD_{it} + \beta_{2it}KDD_{it} + \epsilon_{it} \tag{4}$$

In Section II.B we explain how MO-OLS makes it feasible to estimate models like (4) with fixed effects in intercepts and slopes in large panels. In Section III we use versions of (4) to study the effects of temperature on crop yield across U.S. counties and over time.

Note that our simple model is at the farmer level, while in Section III we will estimate models using county level data. In Appendix A we show that the main predictions of the

[7]Recall I_{it} is a composite of conventional inputs such as capital and labor. Given a common homothetic technology (to rule out scale effects) and common factor prices, all farmers will use inputs in the same proportions. Then r_t can be interpreted as the constant unit price of the optimal bundle of inputs.

FIGURE 1. RELATIONSHIP BETWEEN α^*_{it}, β^*_{2it}, AND KDD_{it}



Note: This graph presents optimal values of β^*_{2it} and α^*_{it} (secondary vertical axis) as a function of KDD_{it}. We use a normalization of p such that $pC_{it}A_tI_{it} = 1$ and $s = -0.01$, which produce values of β^*_{2it} that is within a range consistent with our econometric analysis.

farmer-level model (i.e., that β_{2it} is positively correlated with KDD_{it} and the relationship is approximately log-linear) carry over to a county-level model.

II. Econometric Methods

A. *Previous Approaches to Modelling Crop Yield*

Several recent papers estimate effects of temperature on crop yield using the "degree day" approach. This recognizes that moderate temperatures are beneficial for yield, while high temperatures cause damage. For instance, Schlenker and Roberts (2009) estimate county-level panel-data yield regressions of the following form:

$$(5) \qquad y_{it} = c_i + c_t + \sum_{j=0,3,..}^{39} \beta_j (DD_{j,it} - DD_{j+3,it}) + \beta_{40} PREC_{it} + \beta_{41} PREC^2_{it} + \epsilon_{it}$$

where y_{it} is log yield for county i, year t. The degree day measure $DD_{j,it}$ is total time over the growing season that the crop experiences temperatures above j°C. This model allows the effects of temperature to differ across 3° degree bands. $PREC_{it}$ is total precipitation during the growing season. The c_i and c_t are county and time fixed effects.

Other authors simplify this model by splitting degree days into those above and below 29^oC, which is considered a critical threshold for corn.[8] Cumulative beneficial temperatures ("growing degree days") are given by $GDD_{it} = DD_{0,it} - DD_{29,it}$ and harmful temperatures ("killing degree days") are given by $KDD_{it} = DD_{29,it}$. For instance, Lobell et al. (2011) and Burke and Emerick (2016) estimate equations similar to the following:

$$y_{it} = c_i + c_t + \beta_1 GDD_{it} + \beta_2 KDD_{it} + \beta_3 PREC_{it} + \beta_4 PREC_{it}^2 + \epsilon_{it} \tag{6}$$

This is basically equation (2) of Section I, with county and time effects used to capture the A_t, μ_i and I_{it} terms, and precipitation added. In equations (5) and (6) the key parameter of interest is $\beta_2 < 0$, which captures the extent to which high temperatures reduce crop yield. The county fixed effects c_i capture intercept heterogeneity that is county i specific (such as soil quality), while the time effects c_t capture changes in total factor productivity that are common across counties but vary between years t.

Several studies use the degree day approach to estimate the extent of adaptation to high temperatures, motivated by the idea that the extent of historical adaptation informs us about the scope for future adaptation. For instance, Schlenker and Roberts (2009) test for evidence of historical adaptation to high temperatures by running regressions like (5) after splitting the sample into northern and southern U.S. states and also by 1950-1977 and 1978-2005 periods. Surprisingly, they find coefficients do not differ significantly by region or time, which they take as evidence that historical adaptation has been very limited.

Burke and Emerick (2016) note that, in conventional county fixed-effects models with homogeneous slopes, effects of temperature are identified off of short-run (i.e., annual) deviations of temperature from county means. They adopt a "long difference" approach to try to estimate the response of yields to long-term changes in temperature. Specifically,

[8] At temperatures below 29°C the corn plant can be viewed as a machine for converting heat, water and nutrients into corn. But temperatures above 29°C hamper photosynthesis, predominantly by increasing the need for soil water to sustain carbon assimilation, and by increasing the rate of transpiration (which drains the plant's water supply). Both factors contribute to water stress by increasing the vapour pressure deficit (see Lobell et al. (2013) for details). High temperature can also damage plant tissue directly through heat stress.

they estimate a "long difference" regression similar to:

$$\Delta y_{is} = c_s + \beta_1 \Delta GDD_{is} + \beta_2 \Delta KDD_{is} + \beta_3 \Delta PREC_{is} + \beta_4 \Delta PREC_{is}^2 + \Delta \epsilon_{is} \quad (7)$$

where ΔX_{is} is defined as the change in the average value X in county i of State s from 1998-2002 to 1978-82. Comparing estimates of (7) with a conventional panel-data fixed effects model (6) estimated on annual county-level from 1980 to 2000, they find similar coefficients on KDD. Thus, they infer that adaptation was fairly minor over this period.[9]

It is important to recognize that all the approaches we have discussed rely on fixed-effects models with homogeneous slopes to provide consistent estimates of the sensitivity of crop yield to temperature. Appendix B demonstrates that this assumption is unlikely to hold in practice, as correlation between the (heterogeneous) slope coefficients and temperature, arising due to adaptation, leads to bias in estimating temperature effects.

In contrast to studies we have discussed so far, Butler and Huybers (2013) run separate regressions by county for 1981 to 2008, and conclude from the county-specific coefficients that substantial adaptation has occurred across counties with different climates. In projecting effects of climate change on crop yields, they argue that losses could be halved by adaptation. The use of county-specific regressions avoids the criticism that the coefficients on temperature are identified from short-run variation (i.e., weather rather than climate). Roberts and Schlenker (2012) study whether corn yield has become less sensitive to high temperatures over time, by estimating time-varying regression parameters, and conclude it has not. These approaches are related to ours, in that slope parameters are either county-specific or time-varying. We generalize these approaches by allowing for *both* county and time effects in slopes, thus allowing for adaptation *both* across counties and over time.

Finally, outside of agriculture, Deschênes and Greenstone (2011) model adaptation by interacting current temperature with mean temperature of a region in order to estimate effects of climate change on mortality. Dell et al. (2012) use this approach to estimate effects of climate change on economic growth. Neither paper finds evidence of systematic heterogeneity in the marginal effect that would be indicative of adaptation.

[9]Liang et al. (2017) and Ortiz-Bobea et al. (2018) consider a related question of whether total factor productivity of U.S. agriculture (across multiple crops) has become more sensitive to the climate over time.

B. Two Approaches to Modelling Adaptation

Here we propose two complementary approaches to modelling crop yield that account for potential adaptation to high temperatures. A useful starting point is the conventional two-way fixed effects specification (FE-OLS) that we repeat for convenience:

$$y_{it} = c_i + c_t + \beta_1 GDD_{it} + \beta_2 KDD_{it} + \beta_3 PREC_{it} + \beta_4 PREC_{it}^2 + \epsilon_{it} \tag{8}$$

In Appendix B we show how, given heterogeneity in the KDD coefficient (β_{2it}) induced by adaptation, the FE-OLS estimate of β_2 in (8) is likely to be an upward biased (i.e., toward 0) estimate of the mean KDD coefficient.[10]

In Section I we presented a simple model that rationalized the widely-used specification in (8). When we extended the model to include adaptation, it implied that the effect of high temperatures on crop yield is log-linear function of KDD_{it}. Thus, one way to capture adaptation is the parametric specification:

$$\begin{aligned} y_{it} =& c_i + c_t + \beta_1 GDD_{it} + \beta_{20} KDD_{it} + \beta_{21}(ln(KDD_{it}) * KDD_{it} - KDD_{it}) \\ &+ \beta_3 PREC_{it} + \beta_4 PREC_{it}^2 + \epsilon_{it} \end{aligned} \tag{9}$$

which implies that the marginal effect of KDD_{it} on y_{it} is the log-linear function:

$$\frac{\partial y_{it}}{\partial KDD_{it}} = \hat{\beta}_{20} + \hat{\beta}_{21} ln(KDD_{it}) \tag{10}$$

The model in equation (9) can be simply estimated using FE-OLS. If $\hat{\beta}_{21} > 0$ it implies that the adverse KDD effect is smaller in hotter counties or time periods.[11] But this approach relies on the parametric assumptions in Section I being correct.

Our second - and more novel - approach to estimating adaptation is to estimate a model with both spacial and temporal heterogeneity in the slope coefficients, as in:

$$y_{it} = c_i + c_t + \beta_{1it} GDD_{it} + \beta_{2it} KDD_{it} + \beta_{3it} PREC_{it} + \beta_{4it} PREC_{it}^2 + \epsilon_{it} \tag{11}$$

[10] This is because FE-OLS ignores the correlation between KDD_{it} and β_{2it} generated by adaptation.

[11] Alternatively, we could allow the KDD coefficient to be a log-linear function of a measure of average temperature in a county, as in Butler and Huybers (2013). But this does not capture adaptation over time within counties.

This approach is more flexible, as we do not have to specify a particular form of nonlinearity for the KDD coefficient. Instead, we allow the slope heterogeneity to be correlated with the regressors. Then, by analyzing the distribution of the estimates $\hat{\beta}_{2it}$ post-estimation, we can determine the nature of the relationship between KDD_{it} and $\hat{\beta}_{2it}$ (if any).

Of course, estimating (11) without further restrictions would result in more unknown parameters than data points. To achieve identification we restrict attention to additive heterogeneity across the county/time dimensions, as in:

$$\beta_{kit} = \beta + \lambda_{ki} + \theta_{kt}, \quad k = 1, .., 4 \tag{12}$$

This set up accommodates adaptation across *both* counties and time periods.

The direct way to estimate (11)-(12) is via OLS, where each regressor is interacted with a full set of dummies for each i and t. We call this the 'brute force' approach. In a panel with large N and T this is infeasible as the regressor matrix grows extremely large.

Recall that in large panels it is standard to estimate models with fixed effects in intercepts by de-meaning the data for each unit prior to running OLS. Frisch and Waugh (1933) showed this simple procedure gives the fixed effects estimator. Similarly, one can estimate models with unit fixed effects in both intercepts and slopes by running OLS regressions at the unit level (Pesaran and Smith 1995). But this approach is not possible for models with time effects, which exhaust all degrees of freedom at the unit level. As a result, the literature lacks a computationally practical approach to estimate models with heterogeneous slopes in large panels in the presence of time effects.

Accordingly, in section II.C we present a computationally practical method to estimate models with additive slope heterogeneity over two dimensions that may be correlated with the regressors, as in (11)-(12). We call this the 'mean observation OLS' estimator (MO-OLS), and it is numerically equivalent to the 'brute force' OLS approach.

Importantly, the data variation that identifies the KDD coefficient(s) in (11)-(12), using the MO-OLS approach, is fundamentally different from that in the FE-OLS model (8). In FE-OLS, slopes are identified from the response of yield to idiosyncratic variation in the regressors (i.e., local weather shocks). But, as we show in Appendix E, if KDD has a permanent/transitory structure, and we use MO-OLS, then the λ_i will be identified from the response of yield to permanent differences in counties' climates, while the θ_t will be identified from responses to aggregate time effects in U.S. weather. Thus, MO-OLS can identify long-term adaptation by farmers (in the form of slope heterogeneity) that is driven by county level climate, or by common time effects.

It is important to be clear about the types of adaptation our approach captures. By adaptation we are broadly referring to particular forms of farmer or natural adaptation that leads to nonlinearity in the production function, which we model as heterogeneity across counties/time in the marginal effect of KDD on crop yield. It captures forms of farmer adaptation including adoption of heat resistant seed hybrids, irrigation, improved water retention in the fields, and sowing density. It does not capture other adaptations such as crop switching, changes to the growing season, or land use changes.[12]

Heterogeneity in slope parameters will also capture any nonlinearities in the production function that arise for reasons other than farmer adaptation, such as inherent nonlinearity in the relation between temperature and yield generated by plant biology or local environmental factors. For the purpose of obtaining unbiased projections of the effects of climate change on crop yield, a model should account for both farmer and natural adaptations.

C. The MO-OLS Algorithm

Here we present the mean observation OLS (MO-OLS) procedure that we use to estimate (11)-(12). This is the first computationally feasible panel-data estimator that allows for fixed-effects slope heterogeneity over space and time. Consider the following generic model that includes fixed effects in both intercept and slopes:

$$y_{it} = \boldsymbol{\beta}'_{it}\boldsymbol{x}_{it} + u_{it} \tag{13}$$

for units $i = 1, \cdots, N$ and time periods $t = 1, \cdots, T$, where $\boldsymbol{x}_{it} = (1, x_{1it}, \cdots, x_{Kit})'$ is a $(K+1) \times 1$ vector of regressors, $\boldsymbol{\beta}_{it} = (\beta_{0it}, \beta_{1it}, \cdots, \beta_{Kit})'$ is a $(K+1) \times 1$ vector of coefficients that vary across individuals and over time, and u_{it} is the idiosyncratic error term. Note that $\boldsymbol{x}_{it}$ includes a constant term, which accordingly allows for intercept

[12]However, in Section IV.D we examine heterogeneity in the effects of climate change across counties. This sheds light on the potential for land use changes to mitigate damage.

heterogeneity across i and t. $\boldsymbol{x}_{it}$ may also include lags of the dependent variable or any of the regressors as needed. We assume **A.1**: u_{it} is iid, **A.2**: the regressors are weakly exogenous $E(x_{kis}u_{it}) = 0 \ \forall k$ for $s \geq t$, and **A.3**: $E(u_{it}^2|x_{kit}) < \infty \ \forall k$.

We further assume **A.4**: the coefficient heterogeneity is additively separable, such that $\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_K)'$ is the constant effect across all observations, $\boldsymbol{\lambda}_i = (\lambda_{0i}, \lambda_{1i}, \cdots, \lambda_{Ki})'$ are the individual effects that vary across every unit in the panel, and $\boldsymbol{\theta}_t = (\theta_{0t}, \theta_{1t}, \cdots, \theta_{Kt})'$ are time effects that vary between each time period.

A 'brute force' approach to estimating (13) is to run OLS on a model that includes: (i) dummy variables for each i and t (to capture fixed effects in the intercept), and (ii) a complete set of interaction terms between the regressors and the i and t dummies (to capture unit/time fixed effects in slopes). This is computationally infeasible in medium to large panels, as it involves $(N + T)(K + 1)$ regressors, making it impractical to store and invert $X'X$, or to solve the linear system $(X'X)\beta = X'Y$.[13] See Appendix C for details.

Instead, MO-OLS constructs consistent estimates of $\boldsymbol{\beta}_{it}$ by running a series of feasible regressions and then removing the resulting bias. MO-OLS does this by combining three types of regressions: pooled, i-specific, and t-specific. First, rewrite (13) as:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}$$

$$v_{it} = \boldsymbol{x}_{it}'\boldsymbol{\lambda}_i + \boldsymbol{x}_{it}'\boldsymbol{\theta}_t + u_{it}$$

Consider the pooled OLS estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}' \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it} y_{it} \right) \tag{14}$$

Expanding on y_{it} and simplifying yields:

$$\begin{aligned} \hat{\boldsymbol{\beta}} =& \boldsymbol{\beta} + \boldsymbol{Q}_{xx,NT}^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\lambda}_i \right) \\ &+ \boldsymbol{Q}_{xx,NT}^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\theta}_t \right) + \boldsymbol{Q}_{xx,NT}^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it} u_{it} \right) \end{aligned} \tag{15}$$

[13] In panels with very large N, such as marketing datasets with many products, even models with fixed effects in the intercept alone are computationally daunting, unless one uses the Frisch and Waugh (1933) 'within' transformation.

where $\boldsymbol{Q}_{xx,NT}^{-1} = \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}$. Next, consider the unit-specific regressions:

$$y_{it} = \boldsymbol{x}_{it}'(\boldsymbol{\beta} + \boldsymbol{\lambda}_i) + v_{it}$$

$$v_{it} = \boldsymbol{x}_{it}'\boldsymbol{\theta}_t + u_{it}$$

The unit-specific OLS regressions yield:

$$\hat{\boldsymbol{\beta}}_i = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}y_{it}\right) \tag{16}$$

Expanding on y_{it} and simplifying yields:

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{Q}_{xx,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\theta}_t\right) + \boldsymbol{Q}_{xx,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}u_{it}\right) \tag{17}$$

where $\boldsymbol{Q}_{xx,T}^{-1} = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}$. Next, consider the set of time-specific regressions:

$$y_{it} = \boldsymbol{x}_{it}'(\boldsymbol{\beta} + \boldsymbol{\theta}_t) + v_{it}$$

$$v_{it} = \boldsymbol{x}_{it}'\boldsymbol{\lambda}_i + u_{it}$$

The time-specific OLS regressions yield:

$$\hat{\boldsymbol{\beta}}_t = \boldsymbol{\beta} + \boldsymbol{\theta}_t + \boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\lambda}_i\right) + \boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}u_{it}\right) \tag{18}$$

where $\boldsymbol{Q}_{xx,N}^{-1} = \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right)^{-1}$.

Given the pooled, unit-specific and time-specific regression results, we construct a preliminary estimate of $\boldsymbol{\beta}_{it}$ as follows:

$$\hat{\boldsymbol{\beta}}_{it}^{Prel} = \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\beta}}_t - \hat{\boldsymbol{\beta}} \tag{19}$$

This preliminary estimate is biased, but we can understand the nature of the bias by substituting (15), (17), and (18) into (19) to obtain the expression:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{it}^{Prel} =& \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\beta}}_t - \hat{\boldsymbol{\beta}} = \\
& \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{Q}_{xx,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\theta}_t\right) + \boldsymbol{Q}_{xx,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}u_{it}\right) + \\
& \boldsymbol{\beta} + \boldsymbol{\theta}_t + \boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\lambda}_i\right) + \boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}u_{it}\right) \\
& - \boldsymbol{\beta} - \boldsymbol{Q}_{xx,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\lambda}_i\right) \\
& - \boldsymbol{Q}_{xx,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\theta}_t\right) - \boldsymbol{Q}_{xx,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}u_{it}\right)
\end{aligned} \tag{20}
$$

This can be written more compactly as:

$$
\hat{\boldsymbol{\beta}}_{it}^{Prel} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t + (\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT}) + \left(\boldsymbol{Q}_{xu,N} + \boldsymbol{Q}_{xu,T} - \boldsymbol{Q}_{xu,NT}\right) \tag{21}
$$

where $\boldsymbol{R}_N \equiv \boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\lambda}_i\right)$, $\boldsymbol{R}_{t,NT} \equiv \boldsymbol{Q}_{xx,NT}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\theta}_t\right)$ and $\boldsymbol{Q}_{xu,N} \equiv \boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}u_{it}\right)$, and $\boldsymbol{R}_T$, $\boldsymbol{R}_{i,NT}$, $\boldsymbol{Q}_{xu,T}$ and $\boldsymbol{Q}_{xu,NT}$ are defined similarly.

The expression in (21) can be decomposed into three parts: First, the true observation (i,t) level coefficients $\boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t$, second, the bias term $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$ that arises from correlation between the regressors and the heterogeneity (including the fixed effects in the intercept), and finally terms $(\boldsymbol{Q}_{xu,N} + \boldsymbol{Q}_{xu,T} - \boldsymbol{Q}_{xu,NT})$ involving the errors. The latter vanish asymptotically given our assumptions on the $\boldsymbol{x_{it}}$ and $\boldsymbol{u}_{it}$.

Crucially, the bias term $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$ can be calculated to arbitrary accuracy and eliminated from (21), leaving a consistent estimator of $\boldsymbol{\beta_{it}}$.[14] We now explain the procedure: MO-OLS uses $\hat{\boldsymbol{\beta}}_i$ as a first stage approximation for $\boldsymbol{\lambda}_i$ in $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT})$ to form $\hat{\boldsymbol{R}}_N$ and $\hat{\boldsymbol{R}}_{i,NT}$, and also uses $\hat{\boldsymbol{\beta}}_t$ as a first stage approximation for $\boldsymbol{\theta}_t$ in $(\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$ to form $\hat{\boldsymbol{R}}_T$ and $\hat{\boldsymbol{R}}_{t,NT}$. Substituting the definitions of $\hat{\boldsymbol{\beta}}_i$ and $\hat{\boldsymbol{\beta}}_t$ given by equations (17)

[14]In the special case that the slope heterogeneity is independent of the regressors, the FE-OLS estimator of equation (8), obtained via a two-way within transformation followed by OLS estimation, gives a consistent estimator of the mean coefficient vector $\bar{\boldsymbol{\beta}}$, while the preliminary estimate of $\boldsymbol{\beta_{it}}$ in (20) will give a consistent estimate of the observation-level coefficients. However, as we discuss in Appendix B, we would expect adaptation to generate correlation between the slope heterogeneity and the regressors (in particular, a positive correlation between KDD and β_{2it}). In that case FE-OLS will generally deliver inconsistent estimates of the mean coefficient vector $\bar{\boldsymbol{\beta}}$.

and (18) into the bias term $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$ we obtain:

$$
\begin{aligned}
&(\hat{\boldsymbol{R}}_N - \hat{\boldsymbol{R}}_{i,NT}) + (\hat{\boldsymbol{R}}_T - \hat{\boldsymbol{R}}_{t,NT}) = (\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT}) + \\
&\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{R}_T + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{Q}_{xu,N}\right) + \boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{R}_N + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{Q}_{xu,T}\right) \\
&- \boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{R}_T + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{R}_N + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{Q}_{xu,N} + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{Q}_{xu,T}\right)
\end{aligned} \tag{22}
$$

This expression is equal to the original bias $(\boldsymbol{R}_N - \boldsymbol{R}_{i,NT}) + (\boldsymbol{R}_T - \boldsymbol{R}_{t,NT})$, plus additional bias terms that relate to $\boldsymbol{Q}_{xu}$ (which is $o_p(1)$ under these assumptions) as well as the slope heterogeneity. By subtracting (22) from (21) we eliminate the original bias from $\hat{\boldsymbol{\beta}}_{it}^{Prel}$, while introducing a new bias. Importantly, the new bias must be smaller in magnitude than the original bias. This is stated as Lemma 1 of the Mathematical Appendix, where we prove the result.

We can repeat this process, again using the $\hat{\boldsymbol{R}}_N$, $\hat{\boldsymbol{R}}_{i,NT}$, $\hat{\boldsymbol{R}}_T$, and $\hat{\boldsymbol{R}}_{t,NT}$ to approximate the new bias term in (22). As we show in Lemma 1, this in turn produces a new bias term that is even smaller in magnitude. Thus, this process can be iterated L times to render the bias arbitrarily small, forming the bias removed estimates:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{it} = \hat{\boldsymbol{\beta}}_i + \hat{\boldsymbol{\beta}}_t - \hat{\boldsymbol{\beta}} + \sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(&\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{1,\ell} + \boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{2,\ell} \\
&- \boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{1,\ell} + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{2,\ell}\right)\Bigg)
\end{aligned} \tag{23}
$$

where $\Gamma_{1,\ell} = \boldsymbol{Q}_{xx,T}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{2,\ell-1})$ and $\Gamma_{2,\ell} = \boldsymbol{Q}_{xx,N}^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Gamma_{1,\ell-1})$ when $\ell > 0$, and where $\Gamma_{1,0} = \hat{\boldsymbol{\beta}}_i$ and $\Gamma_{2,0} = \hat{\boldsymbol{\beta}}_t$. This is a Cauchy sequence in ℓ, so a suitable L can be chosen by terminating the sequence once it converges to a desired tolerance. In practice, small values of L are usually adequate. Equation (23) is simple to construct, as it is a function of only the preliminary estimates $(\hat{\boldsymbol{\beta}}_i, \hat{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\beta}})$ and the covariates $\boldsymbol{x}$.

Theorem 1 states consistency of MO-OLS estimates of the observation-level coefficients $\boldsymbol{\beta}_{it}$ as L goes to infinity, and then N and T jointly go to infinity. As MO-OLS is numerically equivalent to 'brute force' OLS, the proof is relegated to the Mathematical Appendix.

Theorem 1: Consistency of $\hat{\boldsymbol{\beta}}_{it}$

For the model in (13), with **A.1-A.4**, if $L \to \infty$ and subsequently $(N, T) \xrightarrow{j} \infty$, then:

$$\hat{\boldsymbol{\beta}}_{it} - \boldsymbol{\beta}_{it} \xrightarrow{p} 0 \tag{24}$$

Proof: See Mathematical Appendix

***Remark*:** Recall that **A.4** imposes that $\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t$. As in OLS, it does not make sense to discuss consistent estimates of the separate components ($\boldsymbol{\beta}$, $\boldsymbol{\lambda}_i$, $\boldsymbol{\theta}_t$) as they are only identified up to location normalizations.

Given consistent estimates of $\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t$, a researcher may study them ex-post as desired. In some cases, a researcher may be interested in the mean coefficient vector $\bar{\boldsymbol{\beta}} = (\boldsymbol{\beta} + E(\boldsymbol{\lambda}_i) + E(\boldsymbol{\lambda}_t))$. Given consistent estimates of the observation-level coefficients, the mean coefficient vector can be estimated by taking a simple average:

$$\hat{\boldsymbol{\beta}}_{MO} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\boldsymbol{\beta}}_{it} \tag{25}$$

We refer to this as the Mean Observation OLS (MO-OLS) estimate, as it averages the observation-level coefficients. The following theorems provide results for consistency and asymptotic normality. The proofs are given in the Mathematical Appendix.

Theorem 2: Consistency of $\hat{\boldsymbol{\beta}}_{MO}$

For the model in (13), with **A.1-A.4**, if $L \to \infty$ and subsequently $(N, T) \xrightarrow{j} \infty$, then:

$$\hat{\boldsymbol{\beta}}_{MO} - \bar{\boldsymbol{\beta}} \xrightarrow{p} 0 \tag{26}$$

Proof: See Mathematical Appendix

Theorem 3: Asymptotic Normality of $\hat{\boldsymbol{\beta}}_{MO}$

For the model in (13), with **A.1-A.4**, if $L \to \infty$ and subsequently $(N, T) \xrightarrow{j} \infty$ such that $N/T \to \chi$ and $0 < \chi < \infty$, then:

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{MO} - \bar{\boldsymbol{\beta}}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{MO}) \tag{27}$$

where $\boldsymbol{\Sigma}_{MO} = Var(\boldsymbol{\lambda}_i) + Var(\boldsymbol{\theta}_t)$. The asymptotic variance can be consistently estimated nonparametrically by:

$$\hat{\mathbf{\Sigma}}_{MO} = \frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}})(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}})' + (\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}})(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}})'\right) \tag{28}$$

where $\hat{\boldsymbol{\beta}}_{\bar{i}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\beta}}_{it}$ and $\hat{\boldsymbol{\beta}}_{\bar{t}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{it}$.[15]

Proof: See Mathematical Appendix

***Remark*:** MO-OLS extends the 'mean group regression' (MG-OLS) of Pesaran and Smith (1995) by providing a consistent estimate of the average effect $\bar{\boldsymbol{\beta}}$ in the presence of time fixed effects. Time fixed effects would render the MG-OLS estimate inconsistent.

In this article, we are not primarily interested in the average effect of weather on crop yield. Rather, we are primarily interested in the observation-level coefficients $\hat{\boldsymbol{\beta}}_{it}$. We examine the distribution of the $\hat{\boldsymbol{\beta}}_{it}$ to determine if they are correlated with KDD_{it}, and to determine if there are trends in the mean of $\hat{\boldsymbol{\beta}}_{it}$ over time. We use these patterns to quantify and understand historical adaptation.

In Appendix D we present Monte Carlo simulation results where we compare the performance of MO-OLS against several traditional panel data estimators in an environment with multidimensional slope heterogeneity that is correlated with the regressors. Our results show that MO-OLS is able to consistently and efficiently estimate the coefficients in this environment at reasonable sample sizes. Our results also reveal how poorly traditional panel data estimators can perform in this environment. Fixed effects and mean group estimators generate biased estimates of average effects, and depending on the underlying structure of the heterogeneity the bias can be very severe.

III. Data and Variable Construction

We use temperature and precipitation data from Schlenker and Roberts (2009). These data contain daily observations on maximum and minimum temperature, and precipitation, on a grid across the continental U.S., from 1950 to 2015.[16] We map the grid-based

[15] Restrictions on the relative rate of convergence of N and T are required due to the small sample time series bias $O(T^{-1})$, noted by Hurwicz (1950) in the case of weakly exogenous regressors (such as a lagged dependent variable), and to prevent an incidental parameter problem. Thus, the estimator is not appropriate for panels with small T.

[16] Schlenker and Roberts subsequently extended the dataset to 2015 after the publication of their paper. They used daily weather observations from stations to construct the dataset. Temperature observations prior to 1950 are available, but it is harder to convert them to a national grid as fewer weather stations were operational.

data onto counties, weighting grid locations by the location of corn production in each county. Using the daily max/min temperatures, we approximate the hours each day that a crop is exposed to one-degree Celsius temperature intervals using a sinusoidal function:

$$DD_C = \begin{cases} 0 & \text{if } C > T_{max} \\ T_{avg} - C & \text{if } C < T_{min} \\ \frac{\left((T_{avg}-C)cos^{-1}(S)+(T_{max}-T_{min})sin(S)/2\right)}{\pi^{-1}} & \text{otherwise} \end{cases} \tag{29}$$

where C is the temperature in Celsius, T_{max} and T_{min} are the daily max/min temperature, $T_{avg} = \frac{T_{max}+T_{min}}{2}$ and $S = cos^{-1}(\frac{2C-T_{max}-T_{min}}{T_{max}-T_{min}})$. This is consistent with the literature.

The threshold that separates GDD from KDD is $29°C$ ($84°F$) for both corn and soybeans, while the minimum GDD threshold is set at $0°C$, following the literature.[17] Using results from (29), we calculate daily degree day values by forming $GDD_{id} = DD_0 - DD_{29}$ and $KDD_{id} = DD_{29}$ for each county i and day d. We aggregate to annual values of GDD_{it} and KDD_{it} by summing over the days of the growing season. For both corn and soybeans we assume this is May 1st to September 30th (in line with the literature).[18] Precipitation is measured as total inches of rain over the growing season.

Our county-level yield data is from the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service. It covers the 1950 to 2015 period. Following the literature, we exclude counties west of the 100th Meridian.[19] We have $N = 2209$ corn growing counties. Some counties lack yield data for all years, giving an unbalanced panel.

IV. Results

A. Conventional Fixed Effects Models (FE-OLS)

Table 1 presents results from corn yield regressions that incorporate county and/or time fixed effects but do not allow for heterogeneous slope coefficients. As in Section II, we refer to these as FE-OLS models. The model in column (1) is similar to ones in the extant

[17] Butler and Huybers (2013) note that while 29 degrees may appear low as a threshold for damaging temperatures, the temperature experienced by the plant itself is higher than the measured air temperature above the crop canopy.
[18] When we let the growing season vary by county and over time it did not meaningfully change the results.
[19] Note: The 100th Meridian separates the Great Plains to the east from the semi-arid lands to the west. The western counties are much more reliant on irrigation. We tried including them but it did not affect our results.

literature (see equation (8), Lobell et al. 2011, Burke and Emerick 2016). The estimated KDD coefficient is -0.0052, implying an additional degree-day above 29°C leads to a decrease in overall corn yield of 0.52 percent. This effect is precisely estimated with standard errors that are clustered at the State level.[20] The model in column (2) restricts time effects to be quadratic trends at the State level (as in Schlenker and Roberts 2009, Roberts and Schlenker 2012). But the estimate of KDD sensitivity is not significantly different. Neither model (1) or (2) includes adaptation.

Table 1—FE-OLS Estimates of the Impacts of Temperature on U.S. Corn Yields

Specification	(1)	(2)	(3)
GDD	0.0002 (0.0001)	0.0003 (0.0001)	0.0004 (0.0001)
KDD	-0.0052 (0.0006)	-0.0054 (0.0006)	-0.0158 (0.0023)
ln(KDD)*KDD - KDD			0.0022 (0.0005)
Precipitation	0.0007 (0.0002)	0.0010 (0.0002)	0.0007 (0.0002)
Precipitation^2 (x 1000)	-0.0006 (0.0002)	-0.0007 (0.0001)	-0.0006 (0.0002)
Constant	2.6755 (0.2564)	2.3552 (0.2509)	2.1977 (0.2399)
Fixed Effects	Cty, Yr	Cty	Cty, Yr
Quad. Time Trend	N/A	State-specific	N/A
R-squared	0.82	0.82	0.83
Obs.	126,373	126,373	126,373

Notes: Results exclude counties west of the 100th Meridian. The sample period is 1950-2015, and $N = 2209$. Models (1)-(3) differ by type of fixed effects and whether the adaptation variable is included. Standard errors are reported in parentheses, and are clustered at the state level.

The model in column (3), which is motivated by the simple theory in Section I, includes as an additional regressor a nonlinear function of KDD designed to capture adaptation to harsh temperatures (see equation (9)). The added regressor is positive and highly

[20]It is also possible to adopt spatial standard errors as in Conley (1999), where the correlation between county errors are assumed to decline with distance. Since doing this leads to less conservative standard errors we report the standard errors that are clustered by state in the results.

significant, implying that as KDD increases the negative marginal effect of KDD on crop yield gets smaller.[21] This may be due to both farmer and natural adaptations, and we argue that to predict future corn yield it is important to capture both.

According to model (3), the *average* marginal effect of KDD is -0.0082. This is significantly more negative than the average marginal effect in specifications (1)-(2), illustrating the bias in standard FE-OLS models that ignore adaptation (see Section II.A). However, while model (3) implies the average marginal effect of KDD on corn yield is 58% more negative than in model (1), it also implies the effect diminishes as KDD increases.

B. Heterogeneous Slope Models (MO-OLS)

In this section we use MO-OLS to estimate a model that allows for both county and time fixed effects in coefficients on temperature and precipitation, as in equations (11) and (12). In contrast to the FE-OLS model in (9), the heterogeneous slope model allows us to model adaptation without imposing a particular functional form *a priori*. We can use the estimates to investigate the nature of the relationship between the slope heterogeneity and KDD, which is informative about the nature of adaptation. Table 2 presents the results. It gives unweighted and weighted means of the estimated slope coefficients (using average crop acreage of each county as weights), as well as other moments.

The unweighted mean coefficient on KDD is -0.0096, implying one extra degree day of temperatures over 29°C causes a 0.96% reduction in crop yield. This effect is slightly larger than the average marginal effect of -0.0082 we obtained Table 1 column (3) when we modelled adaptation parametrically using the nonlinear KDD coefficient. Furthermore, our MO-OLS estimate of the mean KDD coefficient is about 80% more negative than we obtained using the conventional FE-OLS models in Table 1 columns (1)-(2). This illustrates the substantial bias in estimators that ignore slope heterogeneity.

The MO-OLS estimates imply there is substantial heterogeneity in the model coefficients. The standard deviation of the KDD coefficient is 0.0068, with a 90/10 percentile range of -0.0034 to -0.0161. A Hausman-type test of the null of slope homogeneity, which compares the FE-OLS and MO-OLS estimates, gives a test statistic of 189.03 with a p-value of 0.00.

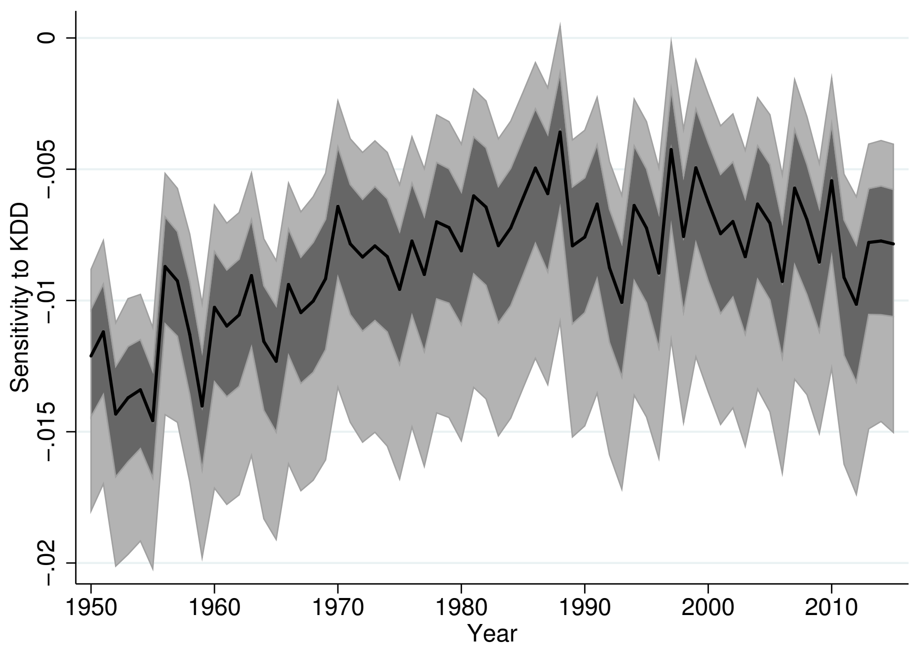
[21] Note that the coefficient on KDD itself becomes significantly more negative at -0.0158 when the nonlinear term is added. The results for the other variables do not change meaningfully from the first two specifications.

Table 2—MO-OLS Estimates of the Impacts of Temperature on U.S. Corn Yields

	Mean	Weighted Mean	Median	Standard Deviation	10th Percentile	90th Percentile
GDD	0.0005 (0.0000)	0.0005	0.0005	0.0005	-0.0001	0.0011
KDD	-0.0096 (0.0003)	-0.0089	-0.0085	0.0068	-0.0161	-0.0034
Precipitation	0.0011 (0.0002)	0.0015	0.0010	0.0028	-0.0020	0.0044
Precipitation^2 (x10,000)	-0.0001 (0.0000)	-0.0001	-0.0001	0.0002	-0.0003	0.0001
Constant	2.7080 (0.1208)	2.8947	2.8225	2.1363	0.1713	4.8419
R-squared	0.88					
Obs.	126,373					

Notes: Results exclude counties west of the 100th Meridian. The sample period is 1950-2015, and $N = 2209$. Standard errors are reported in parentheses.

Figure 2. Distribution of KDD Slope Coefficients across Time and Counties for U.S. Corn



Note: The black line plots the median KDD coefficient from the MO-OLS model in Table 2. The the dark (light) grey areas represent the 25th to 75th, and 10th to 90th percentiles, respectively.

Thus the slope heterogeneity is highly statistically significant.[22] Note that the R^2 improves from .82 to .88 when slope heterogeneity is included in the model.

TABLE 3—ANALYZING TRENDS IN THE MEDIAN KDD COEFFICIENT FOR CORN OVER TIME

Regression Results	β	Std. Err.	β	Std. Err.
$t/10$	0.0008	0.0001	0.0020	0.0002
Constant	-0.0112	0.0005	-0.0133	0.0005
$(t/10)*d_{t>break}$			-0.0022	0.0004
$d_{t>break}$			0.0070	0.0020
Structural Break Test	**Statistic**	**p-value**		
Supremum Wald	42.16	0.00		
Average Wald	27.57	0.00		
Supremum LR	34.24	0.00		
Average LR	24.00	0.00		

Notes: HC3 standard errors are reported for the regression results. The estimated structural break date in the trend and constant is 1989 for corn.

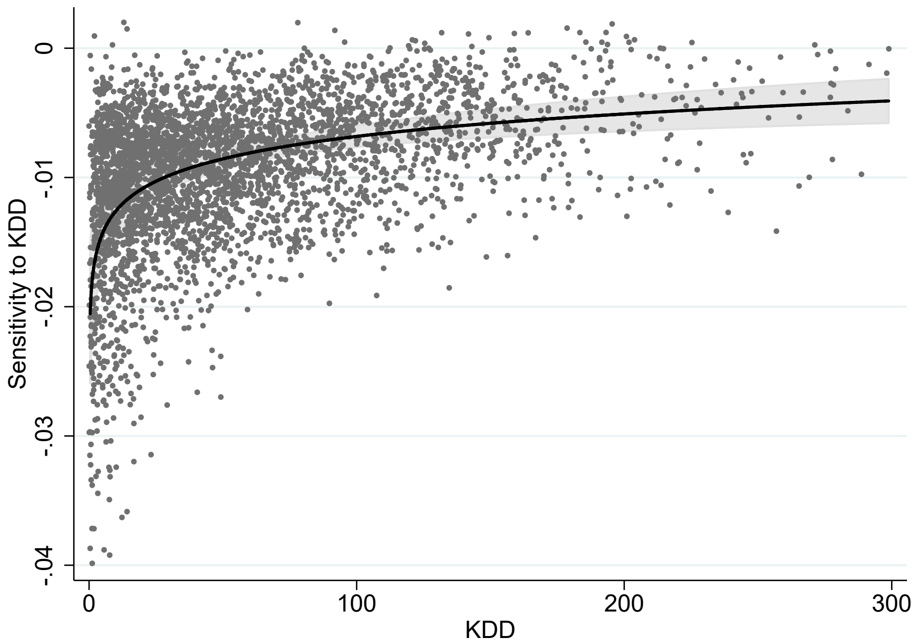
Figure 2 plots the distribution of the KDD coefficients across counties and over time. Clearly, there is significant heterogeneity across counties. The 90/10 percentile range of KDD coefficients is over .010 units in each year. The 25/10 percentile range (i.e., the lower light grey area), is much wider than the 90/75 percentile range, so the distribution has a fat left tail. Figure 2 also shows that the median KDD coefficient follows a clear trend over time. It increases from 1950 until the late 1980s, and then stagnates.

Table 3 presents a regression of the median coefficient for each year on a linear time trend, as well as tests for a structural trend break, using the method of Andrews (1993). The results indicate that a significant break occurred in 1989, coinciding with an extreme drought in the Midwest in 1988-89. This supports the findings of Lobell et al. (2014), who found that while *average* corn yield increased from 1995 to 2010, sensitivity to droughts or high heat increased because of an agronomic trend toward higher sowing densities. This

[22]For this test $H_0 : \boldsymbol{\beta}_{it} = \boldsymbol{\beta} \ \forall \ i, t$ and $H_1 = \boldsymbol{\beta}_{it} \neq \boldsymbol{\beta}$ for a non-zero fraction of the slopes. Pesaran et al. (1996) proposed using the Hausman procedure to test for the presence of slope heterogeneity when $N > T$ using mean group OLS. The same rationale will apply in an environment with multidimensional slope heterogeneity, as FE-OLS remains efficient and consistent under the null and inconsistent under the alternative hypothesis.

may be due to the expansion of crop insurance, see Annan and Schlenker 2015.

Figure 3. Relationship between $\hat{\beta}_{2it}$ and KDD_{it} for U.S. Corn

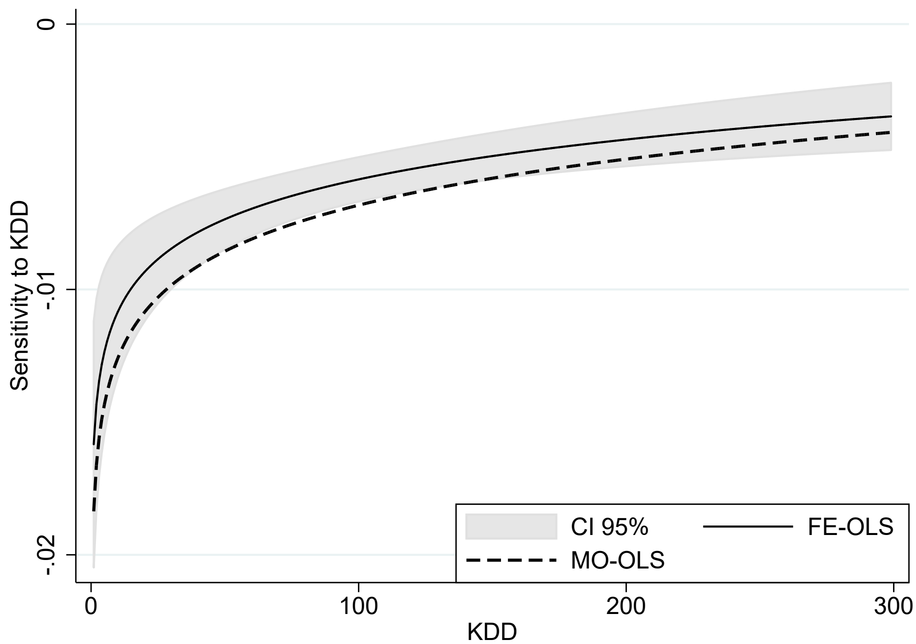


Note: This graph is a scatter plot of a random 3% subsample of the coefficients on KDD_{it} from the MO-OLS model (see Table 2) against KDD_{it} itself. The fitted line was obtained from the regression $\hat{\beta}_{2it} = \alpha_1 + \alpha_2 ln(KDD_{it})$. The estimates are $\alpha_1 = -0.0183$ and $\alpha_2 = 0.0025$ and the 95% confidence interval for the curve is shaded.

Table 3 also presents a regression that accounts for the structural break by including a post-1988 dummy and its interaction with a time trend. This regression implies the time trend was strongly positive up through 1988, after which it turns slightly negative. The median KDD coefficient in 2015 is similar to that in the 1970s, suggesting a lack of sustained progress in adaptation to high temperatures over the last four decades.

The main prediction of the simple model of Section I is that KDD sensitivity should decrease as KDD increases, as farmers have more incentive to adapt. The correlation between $\hat{\beta}_{2it}$ and KDD_{it} is 0.43, supporting this prediction. Figure 3 presents a scatter plot of $\hat{\beta}_{2it}$ and KDD_{it}, as well as a regression of $\hat{\beta}_{2it}$ on $log(KDD_{it})$, which is the best fitting curve to the approximately log-linear relationship. This is remarkably consistent with the simple parametric model of Section I, which also generates a log-linear relation.

FIGURE 4. COMPARISON OF LOG-LINEAR RELATIONSHIPS DERIVED UNDER MO-OLS AND FE-OLS FOR CORN



Note: This graph compares the fitted log-linear relationships between $\hat{\beta}_{2it}$ and KDD_{it} obtained from the county/time specific coefficients estimated with MO-OLS (see Figure 3) and the FE-OLS regression results with adaptation, presented in Table 1 column 3.

In Figure 4 we compare the relationship between $\hat{\beta}_{2it}$ and KDD_{it} that we obtain from the MO-OLS estimated fixed effects vs. the FE-OLS parametric model in Table 1 column (3). The MO-OLS curve implies a more negative marginal effect across all levels of KDD_{it}, but it still lies within the 95 percent confidence interval of the FE-OLS curve.

V. Projecting the Effect of Climate Change on Crop Yield

A. *Forecast Methodology*

Here we project the effects of climate change on U.S. corn yields through to 2100. We present projections using FE-OLS and MO-OLS models with and without adaptation, and we assess the effectiveness of adaptation and emissions reductions in mitigating effects of climate change. We also assess the extent of disagreement in the projections that arises from differences between climate models vs. differences between econometric models.

To project future corn yields, we need predictions of temperature and precipitation for every corn growing county in the U.S. through to 2100. This requires predictions from a climate model, which further requires us to input a greenhouse gas emissions scenario. To assess sensitivity of our predictions to climate model assumptions, we use 19 different climate models and three emissions scenarios. We now describe the procedure in detail:

We utilize weather predictions from 19 general circulation models (GCMs), or simply "climate models," from the Coupled Model Intercomparison Project v5 (CMIP5).[23] Climate models differ in how they represent a number of processes, such as chemical reactions, cloud formation, and vegetation growth. As we will see, they can generate rather different predictions of how the climate will respond to increased greenhouse gas emissions.

Each climate model requires, as an exogenous input, a path for the atmospheric concentration of greenhouse gases. In order to compare the output of climate models consistently, CMIP5 uses four representative concentration pathways: RCP26, RCP45, RCP60, and RCP85.[24] Each RCP embeds assumptions regarding the future trajectory of population growth, technological development and government policies.

We call RCP85 the "business as usual" scenario, as little is done to curb emissions, so atmospheric concentrations of greenhouse gases grow at present rates until 2100. We call RCP45 the "moderate" emissions reduction scenario. It corresponds to policies somewhat more ambitious than current government pledges. Under RCP45, greenhouse gas concentrations climb until 2100, but the growth rate significantly declines after 2060.

Finally, the RCP26 scenario represents "substantial" emissions reductions. It is consistent with the most ambitious targets under the Paris agreement. Greenhouse gas concentrations peak in 2040 and then slowly decline until 2100.[25] This would require significant immediate action, as emissions must start to decline substantially in the very near future.

[23] The core of every GCM is a set of equations that describe the behavior of rotating spheres of continuous fluid that simulate the Earth's atmosphere and oceans. The models we use are: ACCESS 1.0, BNU-ESM, CANESM2, CCSM4, CESM1(CAM5), CSIRO-Mk3.6.0, EC-EARTH, FGOALS-g2, FIO-ESM, GFDL-CM3, GFDL-ESM2G, GISS-E2-R, HadGEM2-ES, IPSL-CM5A-LR, IPSL-CM5A-MR, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, and NorESM1-M. The CMIP protocol was introduced by The World Climate Research Program as part of the Working Group on Coupled Modelling. It ensures the outputs of GCMs are comparable, allowing scientists to analyze them systematically. The fifth version of CMIP is part of the broader effort for the IPCC Fifth Assessment Report.

[24] The RCP scenarios correspond to different radiative forcing values in 2100. For instance, RCP26 results in a forcing value of $+2.6W/m^2$ above pre-industrial levels. We do not consider RCP60 due to its similarity to RCP45.

[25] To be precise, atmospheric concentrations of CO_2 (and all other forcing agents converted to CO_2 equivalence) reach 1240 parts per million by 2100 under RCP85, 575 pp million under RCP45, and 435 ppm under RCP26.

The 19 climate models use greenhouse gas concentrations from the RCP scenarios to produce projections of daily min/max temperature, and precipitation, across a 12x12km grid of the contiguous U.S.. The grid level projections are converted to county level projections using an interpolation procedure is known as "bias-correction and spatial disaggregation" (BCSD). We obtain BCSD output for our 19 climate models from the U.S. Geological Survey Geo Data Portal, who in turn rely on the Bureau of Reclamation (2013).

Finally, we convert the daily temperature and precipitation projections into growing season specific values of GDD_{itrm}, KDD_{itrm}, and $PREC_{itrm}$ for county i, year t, RCP scenario r and climate model m. This is done using the sinusoidal function in (29).

Each climate model was run from 2006 to 2100, so the county-specific projection errors for weather in 2006-15 are observable. We center the projections so the average county-specific projection error of each model for 2006 to 2015 is zero. This corrects for level biases that a particular climate model may have for specific counties, and provides a more accurate picture of the future paths that would be optimally predicted by each model (as optimal predictions should take into account already observed errors). Let $K\tilde{D}D_{itrm}$, $G\tilde{D}D_{itrm}$ and $PR\tilde{E}C_{itrm}$ denote the centered projections of the weather variables.

We are now in a position to use the econometric models in Section IV, paired with each climate model and greenhouse gas emissions scenario, to project future corn yields with and without adaptation. First, we project yields using the conventional FE-OLS model from Table 1 column (1) that does not account for adaptation. This gives:

$$\begin{aligned} \hat{y}_{1itrm} =& 2.6755 + \hat{c}_i + 1.3486 + 0.0002(G\tilde{D}D_{itrm}) - 0.0052(K\tilde{D}D_{itrm}) \\ &+ 0.0007(PR\tilde{E}C_{itrm}) - 0.00000091(PR\tilde{E}C^2_{itrm}) \end{aligned} \tag{30}$$

where all parameters can be read from Table 1, except for $\hat{c}_i$, which is the estimate of the county-specific fixed effect, and 1.3486, which is the 2006-15 mean of the time fixed effect.

Next, we use the FE-OLS specification in Table 1 column (3) that allows for adaptation:

$$\begin{aligned} \hat{y}_{2itrm} =& 2.1977 + \hat{c}_i + 1.3574 + 0.0004(G\tilde{D}D_{itrm}) - 0.0158(K\tilde{D}D_{itrm}) \\ &+ 0.0022(ln(K\tilde{D}D_{itrm}) * K\tilde{D}D_{itrm} - K\tilde{D}D_{itrm}) \\ &+ 0.0007(PR\tilde{E}C_{itrm}) - 0.00000063(PR\tilde{E}C^2_{itrm}) \end{aligned} \tag{31}$$

All parameters are from Table 1, except $\hat{c}_i$ and 1.3574 (the 2006-15 average time effect).

Our MO-OLS model can provide yield projections with or without adaptation. Define $\bar{\hat{\beta}}_{ki}$ as the mean of $\hat{\beta}_{kit}$ over the 2016-15 period for $k = 1, ..., 4$, and similarly for $\bar{\hat{c}}_i$ and $\bar{\hat{c}}_t$. Then, a yield projection that does not allow for future adaptation can be obtained as:

$$\hat{y}_{3itrm} = \hat{c}_i + \bar{\hat{c}}_t + \bar{\hat{\beta}}_{1i} \tilde{GDD}_{itrm} + \bar{\hat{\beta}}_{2i} \tilde{KDD}_{itrm} + \bar{\hat{\beta}}_{3i} \tilde{PREC}_{itrm} + \bar{\hat{\beta}}_{4i} \tilde{PREC}^2_{itrm} \tag{32}$$

where the marginal effect of each variable is fixed at the 2006-2015 mean for each county.[26]

Alternatively, the MO-OLS model can allow for adaptation by setting the coefficient on $\tilde{KDD}_{itrm}$ equal to the predicted value from log-linear relationship shown in Figure 3.[27] That is, we set $\hat{\beta}_{2it} = \alpha_1 + \alpha_2 ln(\tilde{KDD}_{it})$ where $\hat{\alpha}_1 = -0.0183$ and $\hat{\alpha}_2 = 0.0025$:

$$\begin{aligned} \hat{y}_{4itrm} = &\hat{c}_i + \bar{\hat{c}}_t + \bar{\hat{\beta}}_{1i} \tilde{GDD}_{itrm} + (log(\tilde{KDD}_{itrm}) * 0.0025 - 0.0183) \tilde{KDD}_{itrm} \\ &+ \bar{\hat{\beta}}_{3i} \tilde{PREC}_{itrm} + \bar{\hat{\beta}}_{4i} \tilde{PREC}^2_{itrm} \end{aligned} \tag{33}$$

Finally, we aggregate our county-level yield projections to the national level using county average crop areas w_i, as in $\hat{y}_{trm} = \sum_{n=1}^{N} \hat{y}_{itrm} w_i / \sum_{n=1}^{N} w_i$.

B. Projections of Future Crop Yield

Here we present projections of corn yield from 2015 to 2100 using four econometric models, three RCP scenarios and 19 climate models. This gives $4 \cdot 19 \cdot 3 = 228$ scenarios. For each econometric model and emissions scenario r, we report the mean projection across all climate models, $\hat{y}_{tr}$, and 80% prediction intervals around this mean derived from the standard deviation of $\hat{y}_{trm}$ across climate models m. This quantifies the extent of disagreement between climate models, as suggested by Auffhammer et al. (2013).

We present the predictions as percentage changes relative to the 2006-2015 historical average yield. Finally, we apply a five-period moving average to the ensemble average prediction and prediction intervals, simply to reduce noise so as to help visualize trends.

[26] Recall that our estimates capture both farmer and natural adaptation (i.e., inherent nonlinearity in the relation between yield and temperature). By shutting down both, we may exaggerate the impact of farmer adaptation.

[27] There is an asymmetry in (33) in that we allow the KDD coefficient to adapt over time, but we do not let other coefficients change over time. This is because we find no evidence that other parameters adapt to high temperatures in our sample. The correlations of $\hat{c}_{it}$ and $\hat{\beta}_{1it}$ with KDD_{it} are only 0.02 and -0.04, respectively, in sharp contrast to the strong relationship between $\hat{\beta}_{2it}$ and KDD_{it} depicted in Figure 3. It is crucial to recall the distinction between adaptation *per se* (which we can project from historical data) and general technical progress. Predicting changes in c_t and β_{1it} due to general technical progress is a more speculative exercise that we take up in Appendix G.

Appendix F describes the temperature predictions of the 19 climate models. They all predict that *both* KDD and GDD will increase in the absence of emissions reductions. Thus, the overall effect of climate change on yields will be the net balance between the negative effect of higher KDD and the positive effect of higher GDD.

Table 4—Effects of Climate Change on U.S. Corn Yield (pct change)

Year	Conventional FE-OLS	MO-OLS w/o future adaptation	FE-OLS with adaptation	MO-OLS with future adaptation
RCP85				
2030	-04 (-12, 04)	-06 (-16, 04)	-04 (-15, 06)	-06 (-14, 01)
2050	-19 (-34, -05)	-24 (-41, -05)	-21 (-36, 00)	-15 (-24, -05)
2080	-44 (-62, -26)	-51 (-72, -30)	-42 (-59, -26)	-26 (-37, -15)
2100	-62 (-81, -43)	-70 (-89, -51)	-57 (-73, -42)	-36 (-47, -25)
RCP45				
2030	-02 (-12, 09)	-04 (-16, 09)	-02 (-15, 11)	-04 (-13, 04)
2050	-11 (-24, 01)	-15 (-30, 00)	-12 (-26, 01)	-11 (-20, -01)
2080	-21 (-36, -06)	-27 (-44, -09)	-23 (-38, -07)	-16 (-25, -07)
2100	-21 (-38, -04)	-26 (-46, -05)	-22 (-40, 00)	-16 (-26, -06)
RCP26				
2030	-04 (-14, 05)	-07 (-17, 04)	-06 (-16, 05)	-07 (-13, 00)
2050	-07 (-16, 03)	-09 (-20, 01)	-08 (-19, 03)	-08 (-14, -02)
2080	-06 (-15, 04)	-09 (-20, 03)	-07 (-18, 05)	-08 (-15, -01)
2100	-07 (-19, 05)	-10 (-23, 03)	-08 (-21, 05)	-08 (-16, 00)

Notes: Results are expressed in terms of percentage change from the 2006-2015 historical weighted average crop yield. Each number represents the ensemble average over nineteen climate models, while the figures in brackets are the 80% (1.28 standard deviation) prediction intervals.

Table 4 summarizes our main results for corn yield. We report projections from four econometric models: the conventional FE-OLS model in (30), the FE-OLS model with adaptation in (31), and MO-OLS with and without adaptation in (32) and (33). We report results for the three RCP scenarios at four points in time: 2030, 2050, 2080, and 2100. The effect of climate change on corn yield is expressed as a percentage change relative to the 2006-2015 average. We present the ensemble average projection and the 80% (i.e., 1.28 standard deviation) prediction interval (in brackets).

First consider scenarios that ignore adaptation. As we see in Table 4, we predict catastrophic damage to corn yields in the pessimistic RCP85 scenario. The model ensemble average reduction in yield is 70% by 2100 according to the MO-OLS model, and 62% according to the conventional FE-OLS model. As we discussed earlier, the conventional FE-OLS model is likely to understate the impact of climate change because it ignores parameter heterogeneity. Yet we see that it still predicts a very severe impact.

With "moderate" emissions reductions (the RCP45 scenario) we observe smaller, but still significant, reductions in yield, with an ensemble average loss of 26% by 2100 (according to the MO-OLS model). And with "substantial" emissions reductions (as in RCP26, which requires significant immediate action on emissions) we see only a 10% model average reduction in yield (using MO-OLS), and the 80% prediction interval indicates that some climate models even predict no losses.

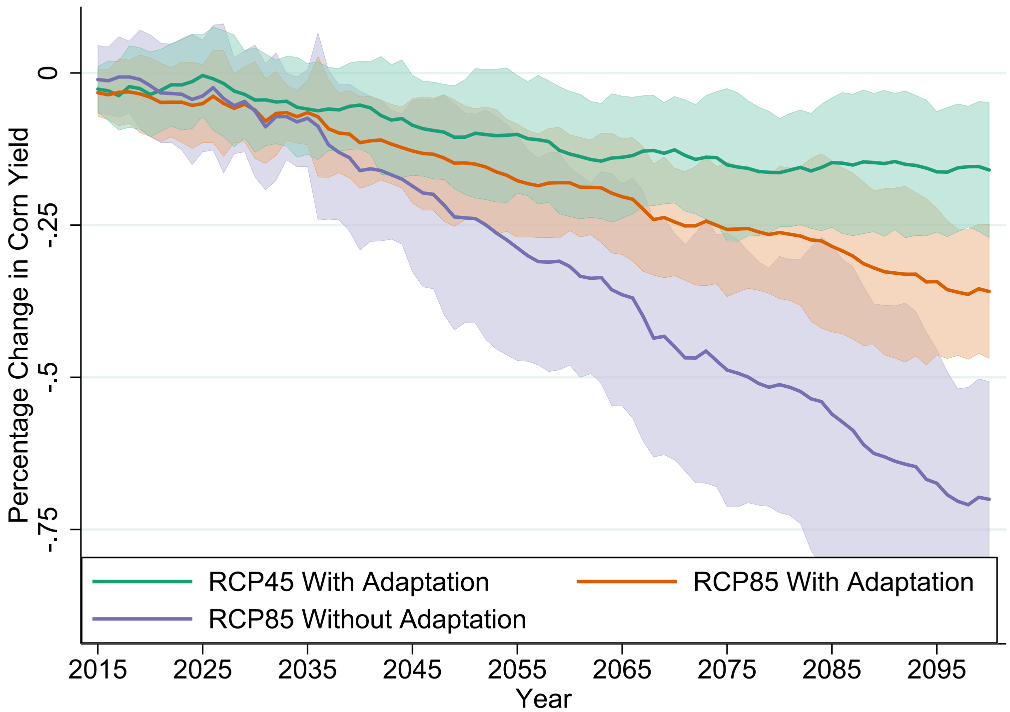
Turning to projections that incorporate adaptation to future climate change, recall that our parametric (FE-OLS) and fixed-effect (MO-OLS) models predict future adaptation based on the historical adaptation patterns shown in Figure 4. The results are reported in the right columns of Table 4. Both models predict adaptation will appreciably mitigate the damage from climate change in the RCP85 scenario. The FE-OLS model predicts adaptation will cause the decline in corn yield in 2100 to drop to 57%. The MO-OLS model implies adaptation is even more effective at mitigating damage: the reduction in crop yield in 2100 decreases from 70% to 'only' 36%.[28]

Of course, with greater emissions reductions KDD increases are smaller, and the scope for adaptation is reduced. For example, given the RCP45 scenario and the MO-OLS model, the mean predicted drop in yield in 2100 is 26% without and 16% with adaptation. And under RCP26 the analogous figures are 10% and 8%, so adaptation is almost irrelevant.

A key takeaway from Table 4 is that adaptation may substantially mitigate damage in the "business as usual" scenario, but damage remains severe. In contrast, even moderate

[28]Thus, the MO-OLS estimates imply greater scope for adaptation to reduce yield loses than the FE-OLS estimates. This is in part because these two approaches model how the effect of KDD_{it} on y_{it} depends on KDD_{it} in different ways. In the MO-OLS approach in (33), the average effect of KDD_{it} is equal to the marginal effect, so adaptation shifts both equally. But in the FE-OLS approach in (31), the average effect of KDD_{it} on y_{it} is given by $\hat{\beta}_{20} + \hat{\beta}_{21}(ln(KDD_{it}) - 1)$, which is always less than the marginal effect in (10). Thus, adaptation in the FE-OLS approach leads to a slower decrease in the average effect relative to the MO-OLS approach. We argue the MO-OLS specification is more intuitive, as it implies that adaptation alters the impact of all units of KDD_{it} on the crop, not just the additional units.

Figure 5. The Effect of Climate Change on Corn Yield by Scenario (MO-OLS)



Note: This graph presents projections of the percentage change in corn yield (relative to the 2006-2015 historical average) for three RCP scenarios, where the solid line is the average projection across nineteen CMIP5 climate models, and the shaded areas are the 80% prediction intervals.

emission reductions are quite effective at reducing damage - e.g., for MO-OLS, compare the 36% mean drop in yield under RCP85 with adaptation with the 26% drop in yield achieved under RCP45 even without adaptation (falling to 16% when adaptation is included). Thus, we find that adaptation alone cannot mitigate most of the damage from climate change - emissions reductions are also required.

Figure 5 presents annual projections of corn yield using the MO-OLS model with and without adaptation. Under RCP85, most climate models predict the decline in yields by 2100 without adaptation will be catastrophic. Adaptation reigns in the most negative projections, and reduces the degree of disagreement between climate models (as is also clear in Table 4), yet the declines in yield remain severe. Under RCP45 with adaptation we predict a moderate mean decline in yield. But there is substantial disagreement between climate models, with the 80% prediction interval ranging from roughly 6% to 30%.

C. Cumulative Losses from 2020-2100

Next we examine the effectiveness of adaptation and emissions reductions at averting yield loss over the whole projection horizon. We calculate total loss of future yield as:

$$Loss_{r,ad} = M^{-1} \sum_{m=1}^{M} \sum_{t=2020}^{2100} (\hat{y}_{trm,ad} - \bar{y}_m) \tag{34}$$

where r is the RCP scenario, ad is the adaptation scenario, and $\bar{y}_m$ is average crop yield from 2006 to 2015 according to the MO-OLS model in the worst-case baseline scenario (i.e., "business as usual" emissions, r=RCP85, and no adaptation, ad=0).[29] The share of damage averted under emissions and adaptation scenario $(RCPn, ad = k)$ for $n = 26, 45, 85$ and $k = 0, 1$ is then defined as $1 - \frac{Loss_{RCPn,ad=k}}{Loss_{RCP85,ad=0}}$.

Table 5—Proportion of Climate Change Damage Averted (pct)

Model	RCP85+Adaptation	RCP45	RCP45+Adaptation	RCP26
MO-OLS	36 (29, 44)	55 (32, 79)	61 (49, 74)	76 (60, 91)
FE-OLS	15 (13, 17)		62 (38, 86)	

Note: Figures are the % reduction in damage relative to the RCP85 scenario with no adaptation using the MO-OLS model.

Table 5 presents the share of damage averted under four scenarios. According to the third column of Table 5, the MO-OLS model predicts a shift from the most pessimistic baseline of RCP85,0 ("business as usual," no adaptation) to RCP45,0 ("moderate reductions," no adaptation) will avert 55% of damage on average, with a 80% prediction interval of 32% to 79%. If we factor in adaptation (RCP45,1) the share of damage averted improves to 61%, and the 80% prediction interval narrows to 49% to 74%. Thus, our point projection is that moderate emissions reductions combined with adaptation will avert more than half the damage to corn yields. But the extent of disagreement across climate models is substantial. Indeed, if we look at the more ambitious RCP26 scenario, the mean damage

[29] We could of course discount future losses in (34), but the proper way to discount losses of future generations is highly controversial. Many have argued against discounting on ethical grounds. In our case, discounting scales down losses, but has almost no effect on relative losses across scenarios, which is our focus.

abatement is 76%, but the 80% prediction interval ranges from only 60% to fully 91%.

According to column (2) of Table 5, the MO-OLS model predicts adaptation alone can avert 36% of total damage to yield (on average),[30] while FE-OLS gives a much smaller figure of only 15%. This is a substantial difference. Yet both econometric approaches agree that adaptation alone (if consistent with historical patterns) cannot avert the majority of the severe damage to yield that is projected to occur with climate change. Significant emissions reductions are necessary to avert substantial reductions in future yields.[31]

D. Distribution of Losses Across U.S. Counties

Here we examine the distribution of losses across counties due to future climate change. Figure 6 graphs the percentage of counties that experience losses as a function of time, RCP, and adaptation scenario. By 2100 the fraction that experience losses approaches one under RCP85 without adaptation. Allowing for adaptation reduces this fraction, but not substantially, as over 80 percent of counties still experience losses by 2100. While not definitive, these results cast doubt on the notion that corn production can be shifted to cooler corn-growing counties as a way to avoid significant damage to yields (at least not U.S. counties that are already producing corn).

Under more ambitious RCP scenarios we still find that the proportion of counties that suffer losses increases over time, but at a much slower rate. In the best case scenario (RCP26 with adaptation) roughly 60% of counties experience losses from about 2040 onward, but about 40% experience gains. This suggests that shifting production to cooler counties may be a more useful strategy if combined with substantial emissions reductions.

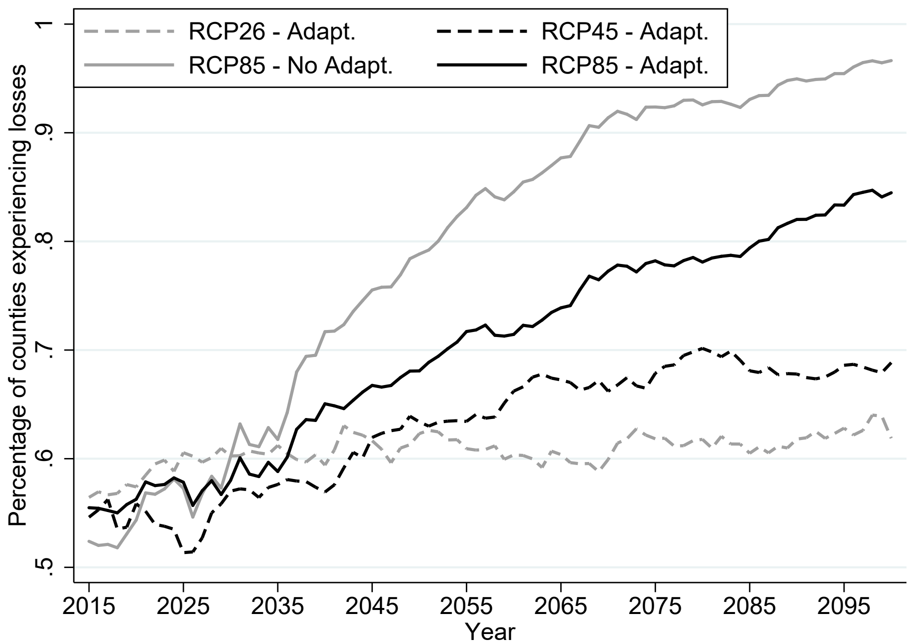
E. Extensions

So far, we have discussed projections holding agricultural technology fixed at current levels (aside from adaptation). In Appendix G we report yield projections that incorporate

[30]As the MO-OLS model predicts a mean yield reduction of 70% in 2100 without adaptation and 36% with adaptation (see Table 4), one might conclude that adaptation averts about half of the damage from climate change. However, the results in Table 5 are based on all years from 2020 to 2100, not merely 2100.

[31]Table 5 does not present FE-OLS results for share of losses averted with emissions reductions but no adaptation. This is because of our earlier finding that the conventional FE-OLS model (without adaptation) seriously understates damage from climate change. Thus we view share-of-loss-averted calculations based on the conventional FE-OLS model as unreliable. In contrast to FE-OLS, the MO-OLS model has the advantage that we can take the estimates from one model and compare results with and without allowing for adaptation.

Figure 6. Proportion of Corn-growing counties experiencing losses from climate change



Note: This graph presents the percentage of corn-crowing counties that experiencing losses from climate change under four combinations of RCP emissions scenario and the MO-OLS model: i) RCP26 with future adaptation, ii) RCP45 with future adaptation, iii) RCP85 without future adaptation, and iv) RCP85 with future adaptation.

projections of technical progress. Specifically, we use VARs to forecast the time effects in the MO-OLS model. We find that under the most optimistic technology forecast combined with RCP26 the mean projection of yield growth keeps pace with population growth. But there is still substantial uncertainty across climate models.

Finally, in Appendix H we present results for soybeans, the second largest U.S. crop. This is an interesting contrast, as soybeans are naturally less sensitive to heat than corn. But the production process for soybeans also exhibits less scope for adaptation (i.e., the KDD coefficient is less heterogeneous, and less highly correlated with KDD).[32] We predict that, without emissions reductions or adaptation, yields will drop by 34% to 74% by 2100. Adaptation is rather ineffective, reducing damage by only 0 to 22%. But emissions reductions are very effective - the moderate RCP45 scenario reduces damage by 36 to 85%.

[32]This is consistent with a relative lack of knowledge on the root architecture and genome of the soybean through most of the sample period (e.g. see Alsajri et al. (2019) and Li et al. (2016)).

VI. Conclusion

We argue that adaptation to high temperatures generates spacial and temporal heterogeneity in the parameters of agricultural production functions. This heterogeneity enters slopes as well as intercepts, and it takes a fixed effects form, as it is correlated with temperature itself. Thus, we propose a new method, that we call "mean observation OLS" (or "MO-OLS") that makes it feasible to estimate panel data models with unit and time fixed effects in both intercepts and slopes in very large panels.

We apply this method to estimate county-level corn yield equations for the U.S. for the 1950-2015 period. We find significant heterogeneity in the effect of temperature on yield, and, as expected, it diminishes as harsh temperatures are more common. This provides strong evidence for historical adaptation between warmer and cooler counties. We also find strong evidence of adaptation over time (i.e., declining heat sensitivity), but this has stalled since 1989. We show that models that do not allow for adaptation produce biased estimates of the effects of high temperatures, and fit the data significantly worse.

We use our econometric model to generate comprehensive county-level projections of corn yields through to 2100, based on future weather scenarios obtained from 19 climate models and three emission growth pathways. Our *econometric* approach is complimentary to work that uses weather predictions from climate models as input into *biological* models of plant growth in order to project future yields (see, e.g., Malcolm et al. 2012).

Our results imply several conclusions regarding the impact of climate change on corn yield: First, absent emissions reductions or adaptation, we predict very severe effects on yield. The average prediction across climate models is -70% by 2100. Second, we predict that almost all corn-growing counties will be adversely affected, implying there is little scope to avert damage by shifting production to cooler regions. Third, we predict that adaptation will avert 36% of damage (on average) in the no emissions reductions scenario. Thus adaptation is important, but can avert less than half of damage to yields.[33] Fourth, on a more optimistic note, we predict that adaptation combined with 'moderate' emissions reductions (i.e., similar to current government pledges) can avert 61% of damage to yields,

[33] An important potential form of adaptation is irrigation. But results in Marshall et al. (2015) suggest that climate change itself will lead to water depletion that will inhibit irrigation in the latter half of this century.

rising to 76% under the more ambitious targets of the Paris agreement. Thus, *plausible* emissions reductions may still avert a large fraction of damage from climate change.

We also attempt to project future technical progress based on past trends (admittedly a rather speculative exercise). We predict that technical progress and adaptation alone (absent emissions reductions) will generate yield growth that lags far behind population growth. But an optimistic projection of technical change, combined with moderate to substantial emissions reductions and adaptation can, together, achieve yield growth roughly in line with population growth according the mean climate model projection. Still, these figures deteriorate quickly under slightly less optimistic technology projections.

A striking feature of our results is the wide variability of projections across climate models. Indeed, we have focused on mean predictions in this conclusion to avoid drowning the reader in a morass of prediction intervals (all of which are presented in detail in the text). Suffice it to say that even our more optimistic emissions/technology/adaptation scenarios put non-negligible mass on rather adverse outcomes. Furthermore, our projections for the second largest U.S. crop, soybeans, are generally a bit more pessimistic. So it is fair to say that climate change poses a substantial *risk* to U.S. agricultural yields, even under the more benign scenarios where our point projections of yield losses are moderate.

REFERENCES

Alsajri, F. A., B. Singh, C. Wijewardana, J. T. Irby, W. Gao, and K. R. Reddy (2019). Evaluating soybean cultivars for low- and high-temperature tolerance during the seedling growth stage. *Agronomy 9*(1).

Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*(4), 821–856.

Annan, F. and W. Schlenker (2015, May). Federal crop insurance and the disincentive to adapt to extreme heat. *American Economic Review 105*(5), 262–66.

Auffhammer, M., S. Hsiang, W. Schlenker, and A. Sobel (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy 7*(2), 181–198.

Burke, M. and K. Emerick (2016). Adaptation to climate change: Evidence from US agriculture. *American Economic Journal: Economic Policy 8*(3), 106–140.

Butler, E. and P. Huybers (2013, January). Adaptation of US maize to temperature variations. *Nature Climate Change 3*, 68–72.

Conley, T. (1999). Gmm estimation with cross sectional dependence. *Journal of Econometrics 92*(1), 1 – 45.

Dell, M., B. Jones, and B. Olken (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics 4*(3), 66–95.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us. *American Economic Journal: Applied Economics 3*(4), 152–185.

Frisch, R. and F. Waugh (1933). Partial time regressions as compared with individual trends. *Econometrica 1*(4), 387–401.

Hsiao, C. (1975). Some estimation methods for a random coefficient model. *Econometrica 43*(2), 305–325.

Hurwicz, L. (1950). Least squares bias in time series. In T. Koopman (Ed.), *Statistical Inference in Dynamic Economic Models*, pp. 365–383. New York: Wiley.

Keane, M. and T. Neal (2018). Comparing deep neural network and econometric approaches to predicting the impact of climate change on agricultural yield. *UNSW Research Working Paper*.

Li, M.-W., D. Xin, Y. Gao, K.-P. Li, K. Fan, N. B. Muoz, W.-S. Yung, and H.-M. Lam (2016, 09). Using genomic information to improve soybean adaptability to climate change. *Journal of Experimental Botany 68*(8), 1823–1834.

Liang, X.-Z., Y. Wu, R. G. Chambers, D. L. Schmoldt, W. Gao, C. Liu, Y.-A. Liu, C. Sun, and J. A. Kennedy (2017). Determining climate effects on us total agricultural productivity. *Proceedings of the National Academy of Sciences 114*(12), E2285–E2292.

Lobell, D., M. Banziger, C. Magorokosho, and B. Vivek (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature climate change 1*, 42–45.

Lobell, D. and M. Burke (2008). Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation. *Environmental Research Letters 3*(3), 1–8.

Lobell, D., G. Hammer, G. McLean, C. Messina, M. Roberts, and W. Schlenker (2013). The critical role of extreme heat for maize production in the united states. *Nature climate change 3*, 497–501.

Lobell, D., M. Roberts, W. Schlenker, N. Braun, B. Little, R. Rejesus, and G. Hammer (2014). Greater sensitivity to drought accompanies maize yield increase in the u.s. midwest. *Science 344*(6183), 516–519.

Malcolm, S., E. Marshall, M. Aillery, P. Heisey, M. Livingston, and K. Day-Rubenstein (2012, July). Agricultural adaptation to a changing climate: Economic and environmental implications vary by u.s. region. Technical Report Economic Research Report 136, United States Department of Agriculture.

Marshall, E., M. Aillery, S. Malcolm, and R. Williams (2015, November). Climate change, water scarcity, and adaptation in the u.s. fieldcrop sector. Technical Report Economic Research Report 201, United States Department of Agriculture.

Ortiz-Bobea, A., E. Knippenberg, and R. G. Chambers (2018). Growing climatic sensitivity of u.s. agriculture linked to technological change and regional specialization. *Science Advances 4*(12).

Pesaran, M., R. Smith, and K. Im (1996). *Dynamic linear models for heterogeneous panels.* Boston: Kluwer Academic.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Porter, J., L. Xie, A. Challinor, K. Cochrane, S. Howden, M. Iqbal, D. Lobell, and M. Travasso (2014). Food security and food production systems. In C. e. a. Field

(Ed.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. pp. 485–533. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Reclamation (2013). Downscaled cmip3 and cmip5 climate projections: Release of downscaled cmip5 climate projections, comparison with preceding information, and summary of user needs. U.S. Department of the Interior, Bureau of Reclamation, Technical Service Center, Denver, Colorado, 116 p., available at: `http://gdo-dcp.ucllnl.org/downscaled_cmip_projections/techmemo/downscaled_climate.pdf`.

Roberts, M. and W. Schlenker (2012). *Is Agricultural Production Becoming More or Less Sensitive to Extreme Heat? Evidence from U.S. Corn and Soybean Yields*. University of Chicago Press.

Schlenker, W. and M. Roberts (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences 106*(37), 15594–15598.

Tannura, M. A., S. H. Irwin, and D. L. Good (2008, February). Weather, technology, and corn and soybean yields in the u.s. corn belt. Technical Report 01, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.

United Nations Population Division (2017). 2017 revision of world population prospects. Can be found online at `https://esa.un.org/unpd/wpp/`.

Wallace, H. A. (1920). Mathematical inquiry into the effect of weather on corn yield in the eight corn belt states. *Monthly Weather Review 48*, 439–446.

Wang, J. C., S. H. Holan, B. Nandram, W. Barboza, C. Toto, and E. Anderson (2012). A bayesian approach to estimating agricultural yield based on multiple repeated surveys. *Journal of Agricultural, Biological, and Environmental Statistics 17*(1), 84–106.

Westcott, P. C. and M. Jewison (2013, July). Weather effects on expected corn and soybean yields. Technical Report FDS-13g-01, U.S. Dept. of Agriculture. Available at: `https://www.ers.usda.gov/webdocs/publications/36651/39297_fds-13g-01.pdf?v=0`.

Mathematical Appendix

Lemma 1

Consider a MxN square matrix $\boldsymbol{B}$ and a Mx1 column vector $\boldsymbol{\omega}$:

$$\boldsymbol{B} = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{pmatrix}, \qquad \boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_m \end{pmatrix}$$

where $b_{m,n} > 0 \; \forall \; m$ and n. Then the Mx1 vector sequence:

$$a_\ell = \begin{cases} \left(\frac{1}{N}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{N}\sum_{n=1}^{N} b_{m,n} a_{\ell-1} & \text{if } \ell \text{ is odd} \\ \left(\frac{1}{M}\sum_{m=1}^{M} b_{m,n}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{m,n} a_{\ell-1} & \text{if } \ell \text{ is even} \end{cases}$$

where $a_0 = \left(\frac{1}{M}\sum_{m=1}^{M} b_{m,n}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{m,n}\omega_m$ is a convergent sequence that has the following pointwise limit in ℓ:

$$\lim_{\ell\to\infty}(a_\ell) = \bar{\boldsymbol{\omega}} = \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\omega_m \quad \textit{pointwise}$$

Proof of Lemma 1

a_0 represents an average of $\boldsymbol{\omega}$ over m for each n which is weighted by $\boldsymbol{B}$:

$$a_0 = \left[\bar{\omega}_0^{n=1}, \bar{\omega}_0^{n=2}, \cdots, \bar{\omega}_0^{n=N}\right]$$

where $\bar{\omega}_0^n = \left(\frac{1}{M}\sum_{m=1}^{M} b_{m,n}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{m,n}\omega_m$. Each element of a_1, in turn, represents a weighted average of all the elements in a_0 over n for each m:

$$a_1 = [\bar{\omega}_1^{m=1}, \bar{\omega}_1^{m=2}, \cdots, \bar{\omega}_1^{m=M}]$$

where $\bar{\omega}_1^m = \left(\frac{1}{N}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \left[b_{m,n=1}\bar{\omega}_0^{n=1} + b_{m,n=2}\bar{\omega}_0^{n=2} + \cdots + b_{m,n=N}\bar{\omega}_0^{n=N}\right]$.

Since $b_{m,n} > 0 \; \forall \; m$ and n, it follows that $inf\{a_0\} \leq inf\{a_1\}$ and $sup\{a_0\} \geq sup\{a_1\}$. If $\exists$ an i, j pair such that $\bar{\omega}_0^{n=i} \neq \bar{\omega}_0^{n=j}$ and $i \neq j$, then it follows that $inf\{a_0\} < inf\{a_1\}$ and

$sup\{a_0\} > sup\{a_1\}$. Only if $\bar{\omega}_0^{n=i} = \bar{\omega}_0^{n=j} \, \forall \, i, j$ will $inf\{a_0\} = inf\{a_1\}$ and $sup\{a_0\} = sup\{a_1\}$. The same argument applies for a_2, which is a weighted average of all elements of a_1 over n for each m, and indeed all subsequent values of ℓ in a_ℓ.

Thus, for every positive real number $\epsilon > 0$ there is a positive integer K such that for all positive integers $i, j > K$, the distance $d(a_i, a_j) < \epsilon$ (i.e. the sequence is convergent).

To demonstrate that $\lim_{\ell \to \infty}(a_\ell) = \bar{\boldsymbol{\omega}}$ pointwise, first note that:

$$sup\{a_0\} \geq \bar{\boldsymbol{\omega}} \geq inf\{a_0\}$$

as it is impossible for:

$$\left(\frac{1}{M}\sum_{m=1}^{M} b_{m,n}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{m,n}\omega_m > \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\omega_m \quad \forall n$$

or:

$$\left(\frac{1}{M}\sum_{m=1}^{M} b_{m,n}\right)^{-1} \frac{1}{M}\sum_{m=1}^{M} b_{m,n}\omega_m < \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\omega_m \quad \forall n$$

when $b_{m,n} > 0 \;\forall\; m$ and n.

Then:

$$sup\{a_1\} \geq \bar{\boldsymbol{\omega}} \geq inf\{a_1\}$$

since it is impossible for:

$$\left(\frac{1}{N}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{N}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n > \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n \quad \forall m$$

or:

$$\left(\frac{1}{N}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{N}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n < \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\right)^{-1} \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} b_{m,n}\bar{\omega}_0^n \quad \forall m$$

when $b_{m,n} > 0\ \forall\ m$ and n, and:

$$\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\bar{\omega}_0^n =$$
$$\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\left[\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\omega_m\right] =$$
$$\left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\right)^{-1}\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}b_{m,n}\omega_m = \bar{\boldsymbol{\omega}}$$

The same argument can be applied to $a_\ell\ \forall \ell > 0$, so that $sup\{a_\ell\} \geq \bar{\boldsymbol{\omega}} \geq inf\{a_\ell\}\ \forall \ell$. Therefore, a_ℓ is a convergent sequence of vectors that always contains within it $\bar{\boldsymbol{\omega}}$, which is accordingly the pointwise limit of the sequence.

Proof of Theorem 1

Given (15), (17), (18), and (23) then:

(35)

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{it} - \boldsymbol{\beta}_{it} = \left(\boldsymbol{Q}_{xu,N} + \boldsymbol{Q}_{xu,T} - \boldsymbol{Q}_{xu,NT}\right) + (-1)^L\Bigg(&\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,L} \\ &+ \boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,L} - \boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,L} + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,L}\right)\Bigg) \\ &+ \sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{1,\ell} + \boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{2,\ell} - \\ &\boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{1,\ell} + \boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{2,\ell}\right)\Bigg)\end{aligned}$$

where $\Theta_{1,\ell} = \boldsymbol{Q}_{xx,T}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,\ell-1})$ and $\Theta_{2,\ell} = \boldsymbol{Q}_{xx,N}^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,\ell-1})$ for $\ell > 0$, $\boldsymbol{\Lambda}_{1,\ell} = \boldsymbol{Q}_{xx,T}^{-1}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\Lambda}_{2,\ell-1})$ and $\boldsymbol{\Lambda}_{2,\ell} = \boldsymbol{Q}_{xx,N}^{-1}(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\boldsymbol{\Lambda}_{1,\ell-1})$ for $\ell > 0$, $\Theta_{1,0} = \boldsymbol{\lambda}_i$, $\Theta_{2,0} = \boldsymbol{\theta}_t$, $\boldsymbol{\Lambda}_{1,0} = \boldsymbol{Q}_{xu,N}$, and finally $\boldsymbol{\Lambda}_{2,0} = \boldsymbol{Q}_{xu,T}$. First, using Lemma

1, (22), and (23) then:

$$\lim_{L\to\infty}\left(\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,L}+\boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,L}\right.$$
$$\left.-\boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,L}+\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,L}\right)\right)=0 \tag{36}$$

To see this, exchange $b_{m,n}$ and ω_m in Lemma 1 for $\boldsymbol{x}_{it}\boldsymbol{x}_{it}'$ and either $\boldsymbol{\lambda}_i$ or $\boldsymbol{\theta}_t$. Since Lemma 1 showed the sequence a_ℓ converges pointwise to $\bar{\boldsymbol{\omega}}$ in ℓ, then also the vector sequence:

$$q_\ell=\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,\ell}+\boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,\ell}$$

must converge to $\boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{1,L}+\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Theta_{2,L}\right)$ in ℓ which gives us the result in (36).

Furthermore, given the Weak Law of Large Numbers, the Continuous Mapping Theorem, and assumptions **A.1**-**A.4**, as $N\to\infty$:

$$\boldsymbol{Q}_{xx,N}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}u_{it}\right)\xrightarrow{p}E(\boldsymbol{Q}_{xx,N}^{-1})E(\boldsymbol{x}_{it}u_{it})=E(\boldsymbol{Q}_{xx,N}^{-1})0=0 \tag{37}$$

Furthermore, as $T\to\infty$:

$$\boldsymbol{Q}_{xx,T}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}u_{it}\right)\xrightarrow{p}E(\boldsymbol{Q}_{xx,T}^{-1})E(\boldsymbol{x}_{it}u_{it})=E(\boldsymbol{Q}_{xx,T}^{-1})0=0 \tag{38}$$

and lastly as $(N,T)\xrightarrow{j}\infty$:

$$\boldsymbol{Q}_{xx,NT}^{-1}\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{x}_{it}u_{it}\right)\xrightarrow{p}E(\boldsymbol{Q}_{xx,NT}^{-1})E(\boldsymbol{x}_{it}u_{it})=E(\boldsymbol{Q}_{xx,NT}^{-1})0=0 \tag{39}$$

Given (37) - (39) and the Continuous Mapping Theorem then:

$$\sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{1,\ell}+\boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{2,\ell}- \boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{1,\ell}+\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{2,\ell}\right)\Bigg)\xrightarrow{p}0 \tag{40}$$

Therefore, as required for Theorem 1:

$$\hat{\boldsymbol{\beta}}_{it}-\boldsymbol{\beta}_{it}\xrightarrow{p}0 \tag{41}$$

Proof of Theorem 2

Since $\bar{\boldsymbol{\beta}}=\boldsymbol{\beta}+E(\boldsymbol{\lambda}_i)+E(\boldsymbol{\theta}_t)=E(\boldsymbol{\beta}_{it})$, $\hat{\boldsymbol{\beta}}_{MO}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{\beta}}_{it}$, and the result from Theorem 1 that $\hat{\boldsymbol{\beta}}_{it}-\boldsymbol{\beta}_{it}\xrightarrow{p}0$ when $L\to\infty$ and then $(N,T)\xrightarrow{j}\infty$, the Weak Law of Large Numbers shows that:

$$\hat{\boldsymbol{\beta}}_{MO}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\boldsymbol{\beta}}_{it}\xrightarrow{p}E(\boldsymbol{\beta}_{it}) \tag{42}$$

which implies Theorem 2.

Proof of Theorem 3

Given (35), (36), (25), and $\bar{\boldsymbol{\beta}}=\boldsymbol{\beta}+E(\boldsymbol{\lambda}_i)+E(\boldsymbol{\theta}_t)$ when $L\to\infty$ then:

(43)

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{MO}-\bar{\boldsymbol{\beta}})=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}((\boldsymbol{\lambda}_i-E(\boldsymbol{\lambda}_i))+(\boldsymbol{\theta}_t-E(\boldsymbol{\theta}_t))+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Psi_{it}+\Xi_{it})$$

where $\boldsymbol{\Psi}_{it}=\left(\boldsymbol{Q}_{xu,N}+\boldsymbol{Q}_{xu,T}-\boldsymbol{Q}_{xu,NT}\right)$ and furthermore $\boldsymbol{\Xi}_{it}=\sum_{\ell=0}^{L}(-1)^{\ell+1}\Bigg(\boldsymbol{Q}_{xx,N}^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{1,\ell}+\boldsymbol{Q}_{xx,T}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{2,\ell}-\boldsymbol{Q}_{xx,NT}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{1,\ell}+\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\Lambda_{2,\ell}\right)\Bigg)$.

Consider now the asymptotics where $(N,T)\xrightarrow{j}\infty$, assumptions **A.1-A.4** and the Weak Law of Large Numbers implies that both $\boldsymbol{\Psi}_{it}\xrightarrow{p}0$ and $\boldsymbol{\Xi}_{it}\xrightarrow{p}0$ (as shown in Theorem 1).

Accordingly:

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}_{MO} - \bar{\boldsymbol{\beta}}) \xrightarrow{d} N\left(0, \Sigma_{MO}\right) \tag{44}$$

where

$$\Sigma_{MO} = Var(\boldsymbol{\lambda}_i) + Var(\boldsymbol{\theta}_t) \tag{45}$$

since $\boldsymbol{\lambda}_i$ and $\boldsymbol{\theta}_t$ are independent by definition.

Now consider the nonparameteric estimate of Σ_{MO} that was proposed in (28):

$$\hat{\boldsymbol{\Sigma}}_{MO} = \frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}})(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}})' + (\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}})(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}})'\right)$$

Given (35) and (36) then:

$$\hat{\boldsymbol{\beta}}_{it} = \boldsymbol{\beta} + \boldsymbol{\lambda}_i + \boldsymbol{\theta}_t + \Psi_{it} + \Xi_{it}, \tag{46}$$

and

$$\left(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}}\right) = \left(\boldsymbol{\lambda}_i - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\lambda}_i\right) + \left(\Psi_{it} - \frac{1}{T}\sum_{t=1}^{T}\Psi_{it}\right) + \left(\Xi_{it} - \frac{1}{T}\sum_{t=1}^{T}\Xi_{it}\right) \xrightarrow{p} (\boldsymbol{\lambda}_i - E(\boldsymbol{\lambda}_i)),$$

and using a symmetric argument

$$\left(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}}\right) \xrightarrow{p} (\boldsymbol{\theta}_t - E(\boldsymbol{\theta}_t))$$

where $\hat{\boldsymbol{\beta}}_{\bar{i}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\beta}}_{it}$ and $\hat{\boldsymbol{\beta}}_{\bar{t}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_{it}$. Therefore:

$$\frac{1}{NT-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left((\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}})(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{t}})' + (\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}})(\hat{\boldsymbol{\beta}}_{it} - \hat{\boldsymbol{\beta}}_{\bar{i}})'\right) \xrightarrow{p} Var(\boldsymbol{\lambda}_i) + Var(\boldsymbol{\theta}_t) \tag{47}$$

and $\hat{\boldsymbol{\Sigma}}_{MO} \xrightarrow{p} \boldsymbol{\Sigma}_{MO}$ as required.