On Model Selection Criteria for Climate Change Impact Studies*

Xiaomeng Cui[†] Dalia Ghanem[‡] Todd Kuffner[§]

June 10, 2021

Abstract

Climate change impact studies inform policymakers on the estimated damages of future climate change on economic, health and other outcomes. In most studies, an annual outcome variable is observed, e.g. agricultural yield, annual mortality or gross domestic product, along with a higher-frequency regressor, e.g. daily temperature. While applied researchers tend to consider multiple models to characterize the relationship between the outcome and the high-frequency regressor, to inform policy a choice between the damage functions implied by the different models has to be made. This paper formalizes the model selection problem in this empirical setting and provides conditions for the consistency of Monte Carlo Cross-validation and generalized information criteria. A simulation study illustrates the theoretical results and points to the relevance of the signal-to-noise ratio for the finite-sample behavior of the model selection criteria. Two empirical applications with starkly different signal-to-noise ratios illustrate the practical implications of the formal analysis on model selection criteria provided in this paper.

*We are grateful to Felix Pretis and Glenn Rudebusch for helpful comments and suggestions.

[†]Jinan University, cuixiaomeng@jnu.edu.cn.

[‡]UC Davis, dghanem@ucdavis.edu.

[§]Washington University, St. Louis, kuffner@wustl.edu.

Keywords: mixed frequency data, climate change, cross-validation, information criteria.

1 Introduction

Using panel data, impacts of climate change have been extensively studied on aggregate economic productivity (Burke et al. 2015; Dell et al. 2012; Hsiang 2010), micro-level productivity and economic returns (Addoum et al. 2020; Deryugina and Hsiang 2017; Somanathan et al. 2021; Zhang et al. 2018), agricultural profits and crop production (Aragón et al. 2021; Burke and Emerick 2016; Cui 2020; Deschênes and Greenstone 2007; Schlenker and Roberts 2009), energy consumption (Auffhammer et al. 2017; Li et al. 2019; Wenz et al. 2017), migration and labor allocation (Cattaneo and Peri 2016; Feng et al. 2010; Jessoe et al. 2018; Mueller et al. 2014), human capital (Garg et al. 2020; Graff Zivin et al. 2018; Park et al. 2020), health and mortality (Barreca et al. 2016; Burke et al. 2018; Deschênes and Greenstone 2011; Heutel et al. 2017), and conflicts (Harari and Ferrara 2018; Hsiang et al. 2011, 2013). While researchers tend to consider multiple models in their analysis of the relationship between the outcome and temperature, both researchers and policymakers have to choose between the implied damage functions, either to use them in an integrated assessment model and/or to inform policy regarding climate mitigation and adaptation. This paper formalizes the model selection problem in climate change impact studies and provides conditions for model selection consistency for a class of model selection criteria in this context.

In typical climate change impact studies, for $i = 1, 2, \dots, n$, $t = 1, 2, \dots, T$, we observe an outcome Y_{it} and a regressor $W_{it\tau}$, which is observed at (higher) frequency, $\tau = 1, 2, \dots, H$. Practitioners tend to present results for a set of models $\{\mathcal{M}_\alpha\}_{\alpha=1}^{A}$, where each model uses different summary statistics of the higher-frequency weather variable as regressors in a fixed effects model, specifically $X_{it,\alpha} \equiv X(\mathcal{W}_{it}, \mu_\alpha) = \mu_\alpha(\mathcal{W}_{it})$, where $\mathcal{W}_{it} \equiv \{W_{it\tau}\}_{\tau=1}^{H}$. Among the most commonly used summary statistics of temperature used in regressions are the annual average (e.g., Dell et al. 2012), various degree day measures (e.g., Burke and Emerick 2016), seasonal averages (e.g., Mendelsohn et al. 1994) as well as temperature bins (e.g., Deschênes and Greenstone 2011). To capture nonlinearities in the annual average temperature, a quadratic function there of is sometimes used (e.g., Burke et al. 2015). For a given α, $\mathcal{M}_\alpha$ specifies a linear model,

$$Y_{it} = X'_{it,\alpha}\beta_\alpha + a_{i,\alpha} + u_{it,\alpha}. \tag{1}$$

Here, $a_{i,\alpha}$ is a fixed effect, $u_{it,\alpha}$ constitues idiosyncratic shocks. The subscript it in $X_{it,\alpha}$ means that, for all α, the regressors are a function of the same $\mathcal{W}_{it}$. This is a key feature of this model selection problem. In practice, additional covariates, year fixed effects and flexible time trends are included. To simplify our presentation, we do not include these additional features. However, our analysis extends in a straighforward manner to accommodating them as we show in our empirical applications.

We first formalize the model selection problem in climate change impact studies. This formal treatment allows us to distinguish it from the classical variable selection problem in linear regression. We note that even though the models considered in this empirical context are linear in the parameters, they all consist of summary statistics of the same underlying high-frequency regressor. This observation has implications for the definitions of nested and non-nested models in this problem. We demonstrate the differences in these definitions relative to the classical variable selection problem using analytical examples.

Next, we examine the conditions under which Monte Carlo cross-validation (MCCV) as well as generalized information criteria (GIC) deliver model selection consistency. Building on the large statistics literature on cross validation (Arlot and Celisse 2010), we present conditions under which Monte Carlo cross-validation (MCCV) can deliver consistent model selection. Consistent with Shao (1993), the MCCV with a vanishing training-to-full sample ratio is shown to be model selection consistent if at least one of the models under consideration nests the true model. Since the formal justification for the consistency of the MCCV relies on the true model being under consideration as well as the homoskedasticity and serially uncorrelatedness of the error term, we also include conditions for consistency of model selection via GICs, which is a general class of model selection criteria that include the

Akaike information criterion (AIC) and the Bayesian information criterion (BIC).[1] This is a natural step given the similarity in asymptotic behavior between AIC (BIC) and MCCV with fixed (vanishing) training-to-full sample ratios (Shao 1997).[2] Furthermore, examining GICs allows us to formally examine the case of misspecification (where none of the models under consideration nests the true model).

Adapting the conditions on model selection consistency of GICs (Sin and White 1996; Vuong 1989), we show that the information criteria proposed in Sin and White (1996), hereinafter SW, are consistent regardless of whether all models under consideration contain the true DGP or not. Consistent with the MCCV results, we find that AIC is not model selection consistent, while BIC is only consistent when at least one of the models under consideration contains the true DGP. However, when this does not hold, the BIC may be inconsistent. This is a novel illustration of the BIC inconsistency property, which was established previously (Hong and Preston 2012; Sin and White 1996). We formally illustrate the special features of our setting that lead to this result despite the fact that all models under consideration are linear in the parameters. The issue specifically stems from the fact that the regressors in the different models are summary statistics of the same underlying high-frequency regressor.

We demonstrate the theoretical results via simulations. Our baseline simulation results confirm our theoretical predictions. When the true DGP is in the set of models under consideration, the MCCV with vanishing training-to-full sample ratios, BIC, and the SW criteria select it with probability approaching one as n increases. AIC and MCCV with fixed training-to-full sample ratios tend to choose less parsimonious models that nest the true DGP. When all models under consideration are misspecified however, only the SW criteria exhibit model selection consistency. These results are consistent with our theoretical results as well as the results in Shao (1997) regarding the relationship between AIC, BIC and MCCV. In order to better understand the finite-sample behavior of the model selection criteria, we

[1]These only require the asymptotic normality of the parameter estimators and hence can accommodate heteroskedasticity, serial and spatial dependence as long as a central limit theorem applies.

[2]It is important to point out that this issue is not well established in the empirical literature (Newell et al. 2021).

examine the behavior of GICs in designs with varying levels of signal-to-noise ratio (SNR) (c.f. Hastie et al. 2020, for example). We find that for low SNR, the SW criteria may choose the null model, even though the true DGP is a nontrivial function of temperature. For these low SNR settings, our simulations indicate that BIC can choose the true DGP with high probability. These simulation results point to the importance of considering different model selection criteria as well as the signal-to-noise ratio in practice.

To illustrate the empirical relevance of our results, we conduct two empirical applications with starkly different signal-to-noise ratios. For each application, we include the models with different damage functions considered in the literature as well as more flexible versions thereof. We first consider the relationship between temperature and crop yields which exhibits sizeable signal-to-noise ratios of about 30-40%. In this application, while different criteria select different models, the associated damage functions are qualitatively and quantitatively similar. Consistent with our theoretical predictions, the SW criteria choose the smallest among those models.

The second application we consider examines the GDP-temperature relationship which suffers from very low signal-to-noise ratios (<1%). In this application, the SW criteria choose the model with no temperature over any other model we consider.[3] While the remaining criteria select models that include temperature, these models deliver qualitatively and quantitatively different damage functions. Importantly, despite the nearly negligable signal-to-noise ratios for all models in this application, some models yield statistically significant and quantitatively large coefficients. Given the inconsistency in these results, we employ smoothing splines, a fully data-dependent procedure, to select the response function. For the unbalanced panel, the smoothing spline fit yields an inverted U-shaped curve, similar to the quadratic model chosen by the BIC, however the magnitude of the estimates is substantially smaller. In the balanced case, the spline is very close to the zero horizontal line. Hence, the smoothing spline estimates are more in line with the null model selected by the SW criteria.

[3] These results are consistent with Newell et al. (2021), who conduct an extensive model comparison for this application allowing not only with different damage functions, but also different controls.

This paper has several practical implications for the use of model selection criteria in climate change impact studies. First, rather than reporting the results of a particular model selection criterion, applied researchers should report the results of the different model selection criteria we consider. The reasoning behind this recommendation stems from the fact that if the true response function is nested in one of the models under considertaion, then the different model selection criteria *should* select models that have similar response functions, albeit with different number of parameters. This is well illustrated in the yield-temperature example. Second, the signal-to-noise ratio should be reported given its implications for the finite-sample performance of the criteria. Finally, for outcomes where the scientific literature and/or economic theory do not specify the response function, a fully data-driven procedure should be used.

This paper builds on the vast literature on the asymptotic properties of model selection criteria (e.g. Arlot and Celisse 2010; Claeskens and Hjort 2008). Section 2.1 reviews the classical variable selection problem in linear regression and relevant results (Shao 1993, 1997). The literature on the asymptotic behavior of GICs in nonlinear model selection problems and particularly the pseudo-inconsistency of BIC are relevant here (Hong and Preston 2012; Sin and White 1996; Vuong 1989). Finally, while the data setting in this problem resembles the mixed data sampling (MiDaS) literature (Andreou and Ghysels 2006; Ghysels et al. 2006, 2007), the modeling objectives in both literatures are distinct. The objective in the MiDaS literature seeks to estimate the differential impact on the outcome of different lags of the high-frequency regressor, whereas the climate change impacts literature seeks to use summary statistics for the high-frequency regressor to characterize its relationship with the outcome of interest.

The paper is organized as follows. Section 2 formalizes the model selection problem in climate change impact studies contrasting it to the classical variable selection problem in linear regression. Section 3 presents conditions for consistent model selection for MCCV and GICs. Sections 4 and 5 provide simulation and empirical illustration of the theoretical results.

2 Model Selection Problem in Climate Change Impact Studies

In this section, we formalize the model selection problem in the climate change impact studies. Since *at first glance* this problem seems akin to a simple variable selection problem in linear regression, we first provide a brief review of this classical problem. Then, we proceed to show how the model selection problem examined here departs from it in Section 2.2.

2.1 Concise Review of Variable Selection in Linear Regression

Before we proceed to examine the model selection problem in the climate change impacts literature, we review the classical variable selection problem in linear regression. Consider a multiple linear regression model with cross-sectional data, $\mathcal{M}_\alpha$, for $i = 1, \ldots, n$,

$$y_i = x'_{i,\alpha}\beta_\alpha + \epsilon_{i,\alpha}, \tag{2}$$

where y_i is the outcome variable, $x_{i,\alpha}$ is a $k_\alpha \times 1$ vector of regressors. Given a set of p regressors, there are 2^p possible combinations of these regressors that one could consider.[4] Model selection criteria allow a practitioner to select between those models.

The properties of MCCV and GICs are well-understood in this context. For systematic treatments of this vast literature, see Shao (1997), Claeskens and Hjort (2008) and Arlot and Celisse (2010). If the true model is finite-dimensional, Shao (1993) shows that the MCCV with vanishing training-to-full sample ratios is model selection consistent, whereas leave-one-out cross-validation can overfit. Shao (1997) provides a general asymptotic framework to compare MCCV and GIC with different tuning parameter choices. Shao (1997) shows that AIC (BIC) and leave-one-out CV (MCCV with vanishing-to-full sample ratio) have similar asymptotic behavior. If there is a "fixed-dimension correct model" (Shao 1997), BIC would be model selection consistent similar to the MCCV with vanishing training-to-testing ratios. AIC, while inconsistent in this case, would be asymptotically loss efficient.[5] If there is no fixed-dimensional correct model, AIC and leave-one-out cross-validation would be model

[4] In best subset selection, one would select the model that minimizes the criterion in question among all 2^p models.

[5] These result suggest a trade-off between model selection consistency and loss efficiency. Yang (2005)

selection consistent, whereas BIC and MCCV with vanishing training-to-full sample ratios would not.

The above results imply that in order to examine model selection consistency, one has to take a stance on whether the true model is finite-dimensional or not. Following the empirical literature on climate change impacts which rely on flexible parametric models, we assume that true model is finite-dimensional. This assumption is further supported by the models used in the scientific literature to characterize the relationship between temperature and yield, human health and other outcomes, which are finite-dimensional models.

We next proceed to examine the model selection problem in the climate change impacts literature distinguishing it from the classical variable selection problem we reviewed here.

2.2 Formalizing Model Selection in Climate Change Impact Studies

The model selection problem faced by empirical researchers in the climate change impacts literature consists of a choice between a finite set of models, $\mathbb{M} = \{\mathcal{M}_\alpha : \alpha = 1, 2, \ldots, A\}$, where $A < \infty$. Each model $\mathcal{M}_\alpha$ is defined by $\{\mu_\alpha, \beta_\alpha, \Xi_\alpha\}$, where we remind the reader that $X(\mathcal{W}_{it}, \mu_\alpha) = \mu_\alpha(\mathcal{W}_{it})$ is the set of summary statistics, β_α is the model-specific regressor coefficient vector, and Ξ_α is the conditional distribution of $Y_{it,\alpha}|\mathcal{W}_i, a_{i,\alpha}$, where $\mathcal{W}_i \equiv \{\mathcal{W}_{i1}, \ldots, \mathcal{W}_{iT}\}$. We let $\mathcal{M}_\star$, with the outcome equation

$$Y_{it} = X'_{it,\star}\beta_\star + a_{i,\star} + u_{it,\star},$$

denote the most parsimonious model that contains the outcome equation of the DGP, i.e. for the true parameter value $\beta_{\star,o}$

$$Y_{it} = X'_{it,\star}\beta_{\star,o} + a_i + u_{it},$$

where a_i and u_{it} are the individual fixed effects and the idiosyncratic shocks of the outcome equation in the DGP. We recognize that the assumptions that the outcome equation is separable in the regressors and the unobservables, a_i and u_{it}, as well as linear in the parameters

shows that for the case where there is a fixed-dimension correct model, one cannot combine the strengths of AIC and BIC, even using adaptive estimation.

are strong. However, we maintain these assumptions to make progress on the problem at hand.

2.2.1 Nested, Strictly Non-nested and Non-nested Overlapping Models

As pointed out above, all models considered contain different summary statistics of the same underlying high-frequency regressor, hence the models are likely to be overlapping. However, we would like to differentiate between different cases of overlapping models. Assume without loss of generality $k_\alpha < k_\gamma$. Let ω denote a realization of $\mathcal{W}_{it}$. For a fixed realization ω, the realizations of $X_{it,\alpha}$ and $X_{it,\gamma}$ are given by $x_{\omega,\alpha} = x(\omega, \mu_\alpha)$ and $x_{\omega,\gamma} = x(\omega, \mu_\gamma)$, respectively. Let $\mathcal{B}_\alpha$ denote the parameter space of β_α and β_α^k the kth element of β_α.

We next provide formal definitions for when two models, $\mathcal{M}_\alpha$ and $\mathcal{M}_\gamma$, are nested, non-nested overlapping or strictly non-nested.

Definition 1. (i) $\mathcal{M}_\alpha$ is nested in $\mathcal{M}_\gamma$ iff $x_{\omega,\alpha} = R_{\alpha,\gamma} x_{\omega,\gamma}$ for all ω, where $R_{\alpha,\gamma}$ is a $k_\alpha \times k_\gamma$ non-random matrix,

(ii) $\mathcal{M}_\alpha$ and $\mathcal{M}_\gamma$ are non-nested, overlapping iff $\mathcal{M}_\gamma$ does not nest $\mathcal{M}_\alpha$, but $x'_{\omega,\alpha}\beta_\alpha = x'_{\omega,\gamma}\beta_\gamma$ for all ω and some $\beta_\alpha \in \mathcal{B}_\alpha$ and $\beta_\gamma \in \mathcal{B}_\gamma$,

(iii) $\mathcal{M}_\alpha$ and $\mathcal{M}_\gamma$ are strictly non-nested iff they are not nested and $x'_{\omega,\alpha}\beta_\alpha \neq x'_{\omega,\gamma}\beta_\gamma$ for all ω, $\beta_\alpha \in \mathcal{B}_\alpha$ and $\beta_\gamma \in \mathcal{B}_\gamma$.

Note that according to (i), a model contains another if the regressors in the latter can be expressed as a linear combination of the regressors in the former. This is different from the typical linear regression framework where a model contains another if the regressors in the latter are a subset of the regressors in the former, i.e. the elements in $R_{\alpha,\gamma}$ can only be zero or one. We illustrate the above definitions with the following example.

Example 1. *(Annual mean, Quarterly Mean and Quadratic in Annual Mean Models)*
Let $\mathcal{M}_\alpha$ denote the annual mean model, with outcome equation

$$Y_{it} = X'_{it,\alpha}\beta_\alpha + a_{i,\alpha} + u_{it,\alpha} \;, \tag{3}$$

where $X_{it,\alpha} = \bar{W}_{it} \equiv \sum_{\tau=1}^{H} W_{it\tau}/H$. *The quarterly mean model uses instead the quarterly means of* $\mathcal{W}_{it}$ *as regressors. Let* Q_q *denote the set of values of* τ *in each quarter* $q = 1, \ldots, 4$. *For a set* A, $|A|$ *denotes its cardinality. In the quarterly mean model,* $X_{it,\gamma} = (\sum_{\tau \in Q_1} W_{it\tau}/|Q_1|, \ldots, \sum_{\tau \in Q_4} W_{it\tau}/|Q_4|)$. *Then* $\mathcal{M}_\gamma$ *prescribes the outcome equation*

$$Y_{it} = X'_{it,\gamma}\beta_\gamma + a_{i,\gamma} + u_{it,\gamma}. \tag{4}$$

Note that $X_{it,\alpha} = R_{\alpha,\gamma}X_{it,\gamma}$, *where*

$$R_{\alpha,\gamma} = \frac{1}{H}\left(\begin{array}{cccc} |Q_1|, & |Q_2|, & |Q_3|, & |Q_4| \end{array} \right).$$

Hence, $\mathcal{M}_\alpha$ *is nested in* $\mathcal{M}_\gamma$.

The quadratic in annual mean model, $\mathcal{M}_\delta$, *uses as regressors* $X_{it,\delta} = (\bar{W}_{it}, \bar{W}_{it}^2)$. *Even though the quadratic in annual mean and the quarterly mean models are not nested, if* $\beta_\delta^2 = 0$ *and* $\beta_\gamma^k = \beta_\gamma^{k'}$ *for* $k \neq k'$, *with* $k, k' \in \{1, 2, 3, 4\}$, *then both models yield the annual mean model given* $\mathcal{W}_{it}$. *Hence, they are overlapping, non-nested.*

2.2.2 Probability Limits of Fixed Effects Estimators

Given that all models considered here use regressors that are functions of different summary statistics of the same time series, we formalize the (pseudo-)true parameter values of the models under consideration. We first introduce the within-demeaning notation for linear fixed effects estimation. For V_{it}, $\widetilde{V}_{it} = V_{it} - \bar{V}_i$, where $\bar{V}_i = \sum_{t=1}^{T} V_{it}/T$. For $\mathcal{M}_\alpha$, the within-transformation is given by

$$\begin{aligned} Y_{it} &= X'_{it,\alpha}\beta_\alpha + a_{i,\alpha} + u_{it,\alpha} \ , \\ \tilde{Y}_{it} &= \tilde{X}'_{it,\alpha}\beta_\alpha + \tilde{u}_{it,\alpha} \ . \end{aligned} \tag{5}$$

The probability limit of the fixed effects estimator of the above model, which we refer to as the pseudo-true parameter vector of $\mathcal{M}_\alpha$, is denoted by β_α^* and is given in (6). Here, we assume that $\{\{Y_{it}, \mathcal{W}_{it}\}_{t=1}^{T}\}_{i=1}^{n}$ are i.i.d. across i and t.[6] We further assume that the

[6]These assumptions are made to simplify the presentation of the probability limits. Clearly, the probability limits can be derived under weaker conditions

estimation problem is sufficiently regular, and we also assume strict exogeneity of the high-frequency regressor, $E[u_{it}|\mathcal{W}_{i1},\ldots,\mathcal{W}_{iT},a_i]=0$. Then

$$\beta_\alpha^* = \operatorname*{plim}_{n\to\infty}\left(\sum_{i=1}^n\sum_{t=1}^T \tilde{X}_{it,\alpha}\tilde{X}_{it,\alpha}'\right)^{-1}\sum_{i=1}^n\sum_{t=1}^T \tilde{X}_{it,\alpha}\tilde{Y}_{it} = \left(E[\tilde{X}_{it,\alpha}\tilde{X}_{it,\alpha}']\right)^{-1}E[\tilde{X}_{it,\alpha}\tilde{X}_{it,\star}']\beta_{\star,o}. \quad (6)$$

Equation (6) is the counterpart of the omitted variable bias formula in this problem. To gain some intuition for (6), consider the case where both $X_{it,\star}$ and $X_{it,\alpha}$ are scalar. Then

$$\beta_\alpha^* = \frac{E[\tilde{X}_{it,\alpha}\tilde{X}_{it,\star}]}{E[\tilde{X}_{it,\alpha}^2]}\beta_{\star,o} = \rho_{*,\alpha}\sqrt{\frac{E[\tilde{X}_{it,\star}^2]}{E[\tilde{X}_{it,\alpha}^2]}}\beta_{\star,o}, \quad (7)$$

where $\rho_{*,\alpha} = E[\tilde{X}_{it,\alpha}\tilde{X}_{it,\star}]/\sqrt{E[\tilde{X}_{it,\alpha}^2]E[\tilde{X}_{it,\star}^2]}$ is the within-correlation coefficient between $X_{it,\alpha}$ and $X_{it,\star}$. Under the assumption that $\beta_{\star,o}$ is non-zero, the sign and the magnitude of $\beta_\alpha^*/\beta_{\star,o}$ will depend on the within-correlation between $X_{it,\alpha}$ and $X_{it,\star}$ as well as the ratio of their variances. If the within-correlation between the two variables is positive, then β_α^* and $\beta_{\star,o}$ will have the same sign, otherwise β_α^* will have the opposite sign of $\beta_{\star,o}$. Suppose that $X_{it,\alpha}$ and $X_{it,\star}$ have equal within-variance, then β_α^* will tend to be smaller in magnitude the weaker the within-correlation between $X_{it,\alpha}$ and $X_{it,\star}$. This example of attenuation bias is similar to the classical measurement error problem. If $X_{it,\alpha}$ has greater within-variance than $X_{it,\star}$, then the attenuation is greater. In general, the sign and relative magnitude of $\beta_\alpha^*/\beta_{\star,o}$ will depend on both $\rho_{*,\alpha}$ and the ratio of the within-variances of $X_{it,\star}$ and $X_{it,\alpha}$.

Returning to the general (non-scalar) case, if $\mathcal{M}_\alpha$ contains $\mathcal{M}_\star$, i.e. $X_{it,\star}=R_{\star,\alpha}X_{it,\alpha}$ for some $R_{\star,\alpha}$, then

$$\beta_\alpha^* = R_{\star,\alpha}'\beta_{\star,o}. \quad (8)$$

For instance, if $\mathcal{M}_\alpha$ is the quarterly mean model, and $\mathcal{M}_\star$ is the annual mean model,

$$\beta_\alpha^* = \begin{pmatrix} \frac{|Q_1|}{H} \\ \vdots \\ \frac{|Q_4|}{H} \end{pmatrix}\beta_{\star,o}. \quad (9)$$

Since $X_{it,\alpha}$ and $X_{it,\star}$ are summary statistics of $\mathcal{W}_{it}$, if $\beta_{\star,o}$ is non-zero, then we expect all elements of β_α^* to be non-zero, unless $R_{\star,\alpha}$ has zero rows. This is different from the standard variable selection problem, where the pseudo-true parameter value for models that contain $\mathcal{M}_\star$ will have zero elements for variables that are not in the DGP in general.

3 Model Selection Consistency in Climate Change Impact Studies

In this section, we formally examine the conditions under which MCCV and GICs are model selection consistent in the context of the climate change impact studies. Consistent with the previous section, the properties of the model selection criteria we examine confirm that the model selection problem here is not simply a variable selection problem in linear regression. The asymptotic behavior of the model selection criteria is specifically similar to its behavior in nonlinear model selection problems.

3.1 Monte Carlo Cross-Validation

Cross-validation is a very popular method in practice, because it directly measures out-of-sample prediction error and seems "model-free". It has been used in Schlenker and Roberts (2009) and Gammans et al. (2017) to justify their model selection choice. In this section, we establish conditions under which Monte Carlo cross-validation (MCCV) yields consistent model selection.

Let $Y_i = (Y_{i1}, \ldots, Y_{iT})$ and $X_i = (X_{i1}, \ldots, X_{iT})$. Given observations $\{Y_i, X_i\}_{i=1}^n$, to compute the MCCV mean squared error, we randomly draw a collection $\mathcal{R}$ of b subsets of $\{1, \ldots, n\}$ with size n_v (test sample size) and select a model $\widehat{\mathcal{M}}_{cv}$ that minimizes, among $\alpha = 1, \ldots, A$, the criterion given by

$$\hat{\Gamma}^{MCCV}_{\alpha,nT} = \frac{1}{n_v T b} \sum_{s \in \mathcal{R}} \|\mathbb{Y}_s - \hat{\mathbb{Y}}_{\alpha,s^c}\|^2. \tag{10}$$

Here $\mathbb{Y}_s = (Y_i')_{i \in s}$ is an $nT \times 1$ vector that vertically stacks Y_i' for all $i \in s$ and $\hat{\mathbb{Y}}_{\alpha,s^c} = \tilde{\mathbb{X}}_{s,\alpha} \hat{\beta}^{s^c}_\alpha$, where $\tilde{\mathbb{X}}_{s,\alpha}$ denotes the within-demeaned version of $\mathbb{X}_{s,\alpha} = (X'_{i,\alpha})_{i \in s}$ and $\hat{\beta}^{s^c}_\alpha$ is the estimator of the parameter vector of $\mathcal{M}_\alpha$ using the training data set $\{Y_i, \mathcal{W}_i\}_{i \in s^c}$, where s^c denotes the complement of s, i.e. the remaining $b-1$ subsets in the collection $\mathcal{R}$ after removing subset s.

To proceed, we need to express the within-demeaned model in matrix form; c.f. Section 2.2.2. For random variables V_{it}, with $i = 1, \ldots, n$, $t = 1, \ldots, T$, define $\widetilde{V}_i = (\widetilde{V}_{i1}, \ldots, \widetilde{V}_{iT})$, and $\widetilde{\mathbb{V}} = (\widetilde{V}_1, \ldots, \widetilde{V}_n)'$. Further, write U_i, $i = 1, \ldots, n$, to denote the error terms in the true DGP, which are assumed to be conditionally mean zero. Then we can express the within-demeaned version of model $\mathcal{M}_\alpha$ in matrix form as

$$\widetilde{\mathbb{Y}} = \widetilde{\mathbb{X}}_\alpha \beta_\alpha + \widetilde{\mathbb{U}}_\alpha.$$

Similarly to Shao (1993), we study the mean squared prediction error (MSPE) of $\mathcal{M}_\alpha$, which is estimated using $\{Y_i, X_i\}_{i=1}^n$, in predicting out-of-sample observations of Y_i, which we will refer to as Z_i. Assume that the conditional variance of the error terms in the true DGP (which are conditionally mean zero) is equal to $E^0[U_i U_i' | \mathcal{W}_i] = \sigma^2 I_T$, and also assume that $\{Y_i, X_i\}$ are i.i.d. across i. The expectation operators $E^0[\cdot|\mathcal{W}_i]$ in the preceding sentence and $E^0[\cdot|\{\mathcal{W}_i\}_{i=1}^n]$ in the following displayed equation refer to the conditional distribution derived from the true joint distribution of the Y_i and $\mathcal{W}_i$. The MSPE of the fitted model $\mathcal{M}_\alpha$ is given by

$$\begin{aligned} \Gamma_{\alpha,nT} &= \frac{1}{nT} E^0 \left[\sum_{i=1}^n \sum_{t=1}^T (\widetilde{Z}_{it} - \widetilde{X}_{it,\alpha}' \hat{\beta}_\alpha)^2 \middle| \{\mathcal{W}_i\}_{i=1}^n \right] \\ &= \frac{T-1}{T}\sigma^2 + \underbrace{\frac{1}{nT}\sigma^2 k_\alpha}_{\text{model dimension}} + \underbrace{\Delta_{\alpha,nT}}_{\text{“misspecification” error}}, \end{aligned}$$

where $\Delta_{\alpha,nT} = \frac{1}{nT} \beta_{\star,o}' \widetilde{\mathbb{X}}_\star' (I_{nT} - P_\alpha) \widetilde{\mathbb{X}}_\star \beta_{\star,o} \geq 0$ and P_α is the projection matrix onto the column space of the design matrix $\widetilde{\mathbb{X}}_\alpha$. The derivation of the above equality is included in Section A of the Appendix for the reader's convenience. Several remarks are in order. The homoskedasticity and serial uncorrelatedness of the idiosyncratic shocks are crucial to obtain a component of the mean squared error that depends on the model dimension. As in Shao (1993), it is convenient to consider two categories of models,

- Category I: $\Delta_{\alpha,nT} > 0$,

- Category II: $\Delta_{\alpha,nT} = 0$, when $X_{it,\star} = R_{\star,\alpha} X_{it,\alpha}$.

The following standard conditions correspond to conditions in Shao (1993) which we have adapted to the fixed effects linear model with stochastic regressors.

Condition 1. *(MCCV Consistency)*

1. *(DGP and Models) For* $i = 1, 2, \ldots, n$, $t = 1, 2, \ldots, T$, $Y_{it} = X_{it,\star}\beta_{\star,\mathrm{o}} + a_i + u_{it}$, *where* $u_{it}|\mathcal{W}_{i1}, \ldots, \mathcal{W}_{iT}, a_i \overset{i.i.d.}{\sim} (0, \sigma^2)$ *across* i *and* t. *For some* $\alpha = 1, \ldots, A$, $\mathcal{M}_\alpha = \mathcal{M}_\star$.
2. *(Model Identifiability)* $\operatorname{plim\,inf}_{n\to\infty} \Delta_{\alpha,nT} > 0$ *for* $\mathcal{M}_\alpha$ *in Category I.*
3. *(Regularity Conditions)*

   i. $\widetilde{\mathbb{X}}'_\alpha \widetilde{\mathbb{X}}_\alpha = O_p(n)$ *and* $\left(\widetilde{\mathbb{X}}'_\alpha \widetilde{\mathbb{X}}_\alpha\right)^{-1} = O_p(n^{-1})$ *for* $\alpha = 1, 2, \ldots, A$,

   ii. $\operatorname{plim}_{n\to\infty} \max_{i\le n, t\le T} w_{it,\alpha} = 0$ $\forall \alpha = 1, 2, \ldots, \mathcal{A}$, *where* $w_{it,\alpha}$ *is the* it^{th} *diagonal element of* P_α,

   iii. $\max_{s\in\mathcal{R}} \left\| \frac{1}{n_v} \sum_{i\in s} \sum_{t=1}^{T} \widetilde{X}_{it,\alpha}\widetilde{X}'_{it,\alpha} - \frac{1}{n_c} \sum_{i\in s} \sum_{t=1}^{T} \widetilde{X}_{it,\alpha}\widetilde{X}'_{it,\alpha} \right\| = o_p(1)$ *for* $\alpha = 1, 2, \ldots, A$.

Condition 1.1 imposes homoskedasticity and serial uncorrelatedness of the idiosyncratic shocks. Condition 1.2 is the model identifiability condition for models in Category I. The regularity condition in 1.3(i) is a high-level condition that ensures the existence of a law of large numbers for $\widetilde{\mathbb{X}}'_\alpha \widetilde{\mathbb{X}}_\alpha / n$, and that it converges to an invertible matrix in probability for any $\alpha = 1, \ldots, A$.

Proposition 1. *Assume Condition 1, and* $n_v/n \to 1$ *and* $n_c = n - n_v \to \infty$, $b^{-1} n_c^{-2} n^2 \to 0$.

(i) If $\mathcal{M}_\alpha$ *is in Category I, then for some* $R_n \ge 0$,

$$\hat{\Gamma}^{MCCV}_{\alpha,nT} = \frac{1}{n_v T b} \sum_{s\in\mathcal{R}} \widetilde{\mathbb{U}}'_s \widetilde{\mathbb{U}}_s + \Delta_{\alpha,nT} + o_p(1) + R_n, \tag{11}$$

where $\widetilde{\mathbb{U}}_s = \widetilde{\mathbb{Y}}_s - \widetilde{\mathbb{X}}_s \beta$.

(ii) If $\mathcal{M}_\alpha$ is in Category II, then

$$\hat{\Gamma}^{MCCV}_{\alpha,nT} = \frac{1}{n_v T b} \sum_{s \in \mathcal{R}} \widetilde{\mathbb{U}}'_s \widetilde{\mathbb{U}}_s + \frac{k_\alpha \sigma^2}{n_c T} + o_p(n_c^{-1}). \tag{12}$$

(iii) It follows that

$$\lim_{n\to\infty} P(\widehat{\mathcal{M}}_{cv} = \mathcal{M}_\star) = 1. \tag{13}$$

The proof is given in Appendix A. The above proposition establishes that if $\mathcal{M}_\star$ is under consideration, then MCCV with $n_v/n \to 1$ and $n_c \to \infty$, hereinafter MCCV-Shao, will select this model with probability tending to one in large samples. Suppose $\mathcal{M}_\star$ is not considered, however some models that contain it (Category II) are in the set of candidate models. Then the above proposition implies that the most parsimonious among those models in Category II will be selected with probability tending to one as $n \to \infty$ by the MCCV-Shao procedure. However, the above does not ensure that if the models considered are all in Category II, i.e. none of the models considered contain $\mathcal{M}_\star$, that the most parsimonious model with the smallest $\lim_{n\to\infty} \Delta_{\alpha,nT}$ will be selected with probability tending to one. We will explore this issue in the simulation section.

In the absence of homoskedasticity and serial uncorrelatedness, it is well-known that it is very difficult to formally justify MCCV. We will, however, examine its performance under weaker conditions in the numerical experiments.

3.2 Generalized Information Criteria

Here we introduce the generalized information criterion (GIC) for this problem. In the linear fixed effects model, we do not need to specify a parametric family for the errors to define the estimator. However, it is computationally convenient to use the result that the linear fixed effects estimator is identical to the conditional maximum likelihood estimator under the additional assumption of Gaussian errors, and the conditioning is on $\bar{Y}_i = \sum_{t=1}^{T} Y_{it}/T$, which is a sufficient statistic for the individual fixed effect (Arellano

2003). To be clear, we do not require the Gaussianity assumption or conditional maximum likelihood estimator for our theoretical results, but we take advantage of this equivalence with the linear fixed effects estimator for convenience. We will use $f(\cdot)$ to denote the relevant density function; its precise meaning should be clear from the context. Let $Y_i = (Y_{i1}, \ldots, Y_{iT})'$ and $X_{i,\alpha} = (X_{i1,\alpha}, \ldots, X_{iT,\alpha})$, the i^{th} contribution to the conditional log-likelihood for $\mathcal{M}_\alpha$ is given by $\log\left(f(Y_i|X_{i,\alpha}, a_{i,\alpha}, \bar{Y}_i; \beta_\alpha, \sigma^2_\alpha)\right) = \log\left(f(\tilde{Y}_i|\tilde{X}_{i,\alpha}; \beta_\alpha, \sigma^2_\alpha)\right) \propto -(T-1)\log(\sigma^2_\alpha) - \sum_{t=1}^{T}\left(\tilde{Y}_{it} - \tilde{X}'_{it,\alpha}\beta_\alpha\right)^2/\sigma^2_\alpha$. The log-likelihood function is hence given by

$$\ell^\alpha_{nT}(\beta_\alpha, \sigma^2_\alpha) = -n(T-1)\log(\sigma^2_\alpha) - \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\tilde{Y}_{it} - \tilde{X}'_{it,\alpha}\beta_\alpha\right)^2}{\sigma^2_\alpha}. \tag{14}$$

In the following, we will work with the conditional profile likelihood function

$$\ell^\alpha_{nT}(\beta_\alpha, \hat{\sigma}^2_\alpha(\beta_\alpha)) \propto -n(T-1)\log\left(\hat{\sigma}^2_\alpha(\beta_\alpha)\right),$$

where

$$\hat{\sigma}^2_\alpha(\beta_\alpha) = \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}(\tilde{Y}_{it} - \tilde{X}'_{it,\alpha}\beta_\alpha)^2 \tag{15}$$

is the constrained maximum likelihood estimator for σ^2_α given a fixed value of β_α. Hereinafter, we let $\hat{\ell}_{nT} \equiv \ell^\alpha_{nT}(\hat{\beta}_\alpha, \hat{\sigma}^2_\alpha(\hat{\beta}_\alpha)) = -n(T-1)\log(\hat{\sigma}^2_\alpha(\hat{\beta}_\alpha))$, where $\hat{\beta}_\alpha = \arg\max_{\beta\in\mathcal{B}} \ell^\alpha_{nT}$ $(\beta_\alpha, \hat{\sigma}^2(\beta_\alpha))$.

The generalized information criterion (GIC) is given by the following

$$GIC_{\alpha,\lambda_{nT}} = \hat{\ell}^\alpha_{nT} - \lambda_{nT}k_\alpha. \tag{16}$$

The term λ_{nT} penalizes model dimension. The choice of $\lambda_{nT} = 2$ corresponds to the AIC, whereas the choice $\lambda_{nT} = \log(nT)$ corresponds to the BIC. One of the attractive features of information criteria is that we can formally justify their behavior under heteroskedasticity, spatial and/or time series dependence by viewing them as a misspecfication of the above log-likelihood. Since we will deal with misspecification in this section, we introduce another definition, which is pseudo-consistency of a model selection procedure following Sin and White (1996).

Definition 2. (Pseudo-Consistency of Model Selection) Let $\mathbb{M} = \{\mathcal{M}_\alpha\}_{\alpha=1}^A$ and $\mathcal{M}_\star$ is not nested in $\mathcal{M}_\alpha$ for any $\alpha = 1, \ldots, A$. Then a model selection criterion $\mathbf{C}$ is said to be pseudo-consistent if $\lim_{n\to\infty} P(\widehat{M}_{\mathbf{C}} = \mathcal{M}_{\mathbf{P}}) = 1$, where $\widehat{M}_{\mathbf{C}}$ is the model selected by criterion $\mathbf{C}$ and $\mathcal{M}_{\mathbf{P}}$ is the most parsimonious model with the smallest Kullback-Leibler divergence from the true data-generating distribution among all models in $\mathbb{M}$.

Using previous results on the behavior of the quasi-log-likelihood ratio statistic (Sin and White 1996; Vuong 1989), we can establish conditions for GIC's consistency and pseudo-consistency. Without loss of generality, consider the choice between two models $\mathcal{M}_\alpha$ and $\mathcal{M}_\gamma$. Assume $k_\alpha < k_\gamma$. Let $LR_{nT}^{\alpha,\gamma} = \hat{\ell}_{nT}^{\alpha} - \hat{\ell}_{nT}^{\gamma}$. Further, write $\hat{\mathcal{M}}_{\lambda_{nT}}$ to denote the model that minimizes the GIC given λ_{nT},

$$P(\hat{\mathcal{M}}_{\lambda_{nT}} = \mathcal{M}_\alpha) = P\left(GIC_{\alpha,\lambda_{nT}} > GIC_{\alpha,\lambda_{nT}})\right) = P\left(LR_{nT}^{\alpha,\gamma} > \lambda_{nT}(k_\alpha - k_\gamma)\right),$$

where $LR_{nT}^{\alpha,\gamma} = \hat{\ell}_{nT}^{\alpha} - \hat{\ell}_{nT}^{\gamma}$. Vuong (1989) establishes that the rate of convergence of the quasi-likelihood ratio statistic under the null hypothesis differs depending on whether the conditional densities under $\mathcal{M}_\alpha$ and $\mathcal{M}_\gamma$ agree at the pseudo-true parameter values or not. In our setting, this is determined by whether the predicted values of the outcome at the pseudo-true parameters of the two models differ or coincide. The following proposition formally states how this applies to our setting. First, we impose a high-level condition for the result.

Condition 2. *(Joint Asymptotic Normality of Estimators)*

$$\sqrt{n}\begin{pmatrix} \hat{\beta}_\alpha - \beta_\alpha^* \\ \hat{\beta}_\gamma - \beta_\gamma^* \end{pmatrix} \xrightarrow{d} N(0, \Sigma), \ \ n \to \infty.$$

Here the mean of the multivariate normal is a $(k_\alpha + k_\gamma) \times 1$ zero vector, and Σ is a $(k_\alpha + k_\gamma) \times (k_\alpha + k_\gamma)$ matrix. Primitive conditions that satisfy the above condition include appropriate assumptions on dependence and moments of the outcome and regressors that ensure the existence of laws of large numbers as well as central limit theorems.

The following proposition gives sufficient conditions for $GIC_{\lambda_{nT}}$ to deliver (pseudo-) consistent model selection in our problem when considering three possible cases with two models. The first two cases are from Vuong (1989, Theorem 3.3) when both models are equal in terms of Kullback-Leibler divergence from the true data-generating distribution. In both cases, a (pseudo-) consistent GIC should select the more parsimonious model. The third case is when one model is strictly better in terms of Kullback-Leibler divergence, in which case this model should be chosen by a (pseudo-) consistent GIC. Recall that $x_{\omega,\alpha}$ is a realization of $X_{it,\alpha}$ given a particular realization of $\mathcal{W}_{it}$. Let $\tilde{\boldsymbol{x}}_{\{\omega_t\}_{t=1}^T,\alpha} \equiv \{\tilde{x}_{\omega_t,\alpha}\}_{t=1}^T$ denote the within-demeaned version of $\boldsymbol{x}_{\{\omega_t\}_{t=1}^T,\alpha} = \{x_{\omega_t,\alpha}\}_{t=1}^T$ given T realizations of $\mathcal{W}_{it}$, i.e. $\{\omega_t\}_{t=1}^T$. Let $f(\tilde{Y}_i|\tilde{X}_{i,\alpha};\beta_\alpha^*) = f(\tilde{Y}_i|\tilde{X}_{i,\alpha};\beta_\alpha^*,\sigma_\alpha^*(\beta_\alpha^*))$, where $\sigma_\alpha^*(\beta_\alpha^*) = \text{plim}_{n\to\infty}\hat{\sigma}_\alpha(\beta_\alpha^*)$. Following Vuong (1989), $E^0[.]$ denotes the expectation with respect to the true joint distribution of Y_i and $\mathcal{W}_i$.

Proposition 2. *Assume Condition 2 holds. The following statements hold as $n \to \infty$.*

1. *Suppose* $E^0[\log(f(\tilde{Y}_i|\tilde{X}_{i,\alpha};\beta_\alpha^*))] = E^0[\log(f(\tilde{Y}_i|\tilde{X}_{i,\gamma};\beta_\gamma^*))]$ *and* $f(.|\tilde{\boldsymbol{x}}_{.,\alpha};\beta_\alpha^*) = f(.|\tilde{\boldsymbol{x}}_{.,\gamma};\beta_\gamma^*)$ *hold. Then*

$$P(\hat{\mathcal{M}}_{\lambda_{nT}} = \mathcal{M}_\alpha) = P\left(GIC_{\alpha,\lambda_{nT}} > GIC_{\gamma,\lambda_{nT}}\right) = P\left(LR_n^{\alpha,\gamma} > \lambda_{nT}(k_\alpha - k_\gamma)\right) \to 1,$$

   if $\lambda_{nT} \to \infty$.

2. *Suppose* $E^0[\log(f(\tilde{Y}_i|\tilde{X}_{i,\alpha};\beta_\alpha^*))] = E^0[\log(f(\tilde{Y}_i|\tilde{X}_{i,\gamma};\beta_\gamma^*))]$ *and* $f(.|\tilde{\boldsymbol{x}}_{.,\alpha};\beta_\alpha^*) \neq f(.|\tilde{\boldsymbol{x}}_{.,\gamma};\beta_\gamma^*)$ *hold. Then*

$$P(\hat{\mathcal{M}}_{\lambda_{nT}} = \mathcal{M}_\alpha) = (GIC_{\alpha,\lambda_{nT}} > GIC_{\gamma,\lambda_{nT}}) = P\left(\frac{1}{\sqrt{nT}}LR_{nT}^{\alpha,\gamma} > \frac{\lambda_{nT}}{\sqrt{nT}}(k_\alpha - k_\gamma)\right) \to 1,$$

   if $\lambda_{nT}/\sqrt{nT} \to \infty$.

3. *Suppose, without loss of generality, that* $E^0[\log(f(\tilde{Y}_i|\tilde{X}_{i,\alpha};\beta_\alpha^*))] > E^0[\log(f(\tilde{Y}_i|\tilde{X}_{i,\gamma};\beta_\gamma^*))]$ *holds. Then*

$$P(\hat{\mathcal{M}}_{\lambda_{nT}} = \mathcal{M}_\alpha) = P(GIC_{\alpha,\lambda_{nT}} > GIC_{\gamma,\lambda_{nT}}) = P\left(\frac{1}{nT}LR_{nT}^{\alpha,\gamma} > \frac{\lambda_{nT}}{nT}(k_\alpha - k_\gamma)\right) \to 1,$$

   if $\lambda_{nT}/(nT) \to 0$.

The proof of the above proposition is immediate from Theorem 3.3 in Vuong (1989). The result is a special case of what has been shown in Sin and White (1996) and Hong and Preston (2012). The above lemma shows that for GIC to be (pseudo-) consistent in all cases, then λ_{nT} has to fulfill three conditions as $n \to \infty$: (a) $\lambda_{nT} \to \infty$, (b) $\lambda_{nT}/\sqrt{nT} \to \infty$, (c) $\lambda_{nT}/(nT) \to 0$. These conditions are satisfied for $\lambda_{nT} = \sqrt{nT \log(\log(nT))}$ or $\lambda_{nT} = \sqrt{nT \log(nT)}$ proposed in Sin and White (1996). However, $BIC_\alpha = GIC_{\alpha,\log(nT)}$ only satisfies (a) and (c), but not (b), which is required for consistency of model selection in Case 2. This pseudo-inconsistency of BIC occurs when $f(.|\tilde{\boldsymbol{x}}_{\cdot,\alpha};\beta^*_\alpha) \neq f(.|\tilde{\boldsymbol{x}}_{\cdot,\gamma};\beta^*_\gamma)$, which is determined by the inequality of the predicted values at the pseudo-true parameters.

In the following section, we demonstrate that if the models considered contain the true DGP, then the predicted values at the pseudo-true parameters are equal given $\mathcal{W}_i$ and hence we expect BIC to be consistent. However, if model selection is conducted among models that do not contain $\mathcal{M}_\star$, i.e. all the models under consideration are misspecified, then this issue may occur and BIC may therefore be pseudo-inconsistent.

3.2.1 Predicted Values at the Pseudo-True Parameters

Let $\tilde{Y}^*_{it,\alpha}(\mathcal{W}_i) \equiv \tilde{X}'_{it,\alpha}\beta^*_\alpha$ denote the within-demeaned predicted value of the outcome for individual i in period t given $\mathcal{W}_i$ using the pseudo-true parameter vector of $\mathcal{M}_\alpha$. Consider two models, $\mathcal{M}_\alpha$ and $\mathcal{M}_\gamma$ where both models contain $\mathcal{M}_\star$, i.e. $\tilde{X}_{it,\star} = R_{\star,\alpha}\tilde{X}_{it,\alpha} = R_{\star,\gamma}\tilde{X}_{it,\gamma}$. By the results above – recall that we established in (8) that $\beta^*_\alpha = R'_{\star,\alpha}\beta_{\star,o}$ for $\mathcal{M}_\alpha$ when $\mathcal{M}_\star$ is nested in it – it follows that

$$\begin{aligned}\tilde{Y}^*_{it,\alpha}(\mathcal{W}_i) &= \tilde{X}'_{it,\alpha}\beta^*_\alpha = \tilde{X}'_{it,\alpha}R'_{\star,\alpha}\beta_{\star,o} = \tilde{X}'_{it,\star}\beta_{\star,o},\\ \tilde{Y}^*_{it,\gamma}(\mathcal{W}_i) &= \tilde{X}'_{it,\gamma}\beta^*_\gamma = \tilde{X}'_{it,\gamma}R'_{\star,\gamma}\beta_{\star,o} = \tilde{X}'_{it,\star}\beta_{\star,o}.\end{aligned} \tag{17}$$

Hence, in this case, both models yield identical predictions given $\mathcal{W}_i$ using their respective pseudo-true parameter vectors. This result holds regardless of the relationship between the two models as long as $\mathcal{M}_\star$ is nested in both of them.

Note that if $\mathcal{M}_\alpha$ is nested in $\mathcal{M}_\gamma$, but the DGP is not contained in either model, they may still have different predictions using their respective pseudo-true parameter vectors. To see this, consider

$$\tilde{Y}^*_{it,\gamma}(\mathcal{W}_i) = \tilde{X}'_{it,\gamma}\beta^*_\gamma = \tilde{X}'_{it,\gamma}\left(E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\gamma}]\right)^{-1} E[\tilde{X}_{it,\gamma}X'_{it,\star}]\beta_{\star,o}\ , \tag{18}$$

$$\begin{aligned}\tilde{Y}^*_{it,\alpha}(\mathcal{W}_i) &= \tilde{X}'_{it,\alpha}\beta^*_\alpha = \tilde{X}'_{it,\alpha}\left(E[\tilde{X}_{it,\alpha}\tilde{X}'_{it,\alpha}]\right)^{-1} E[\tilde{X}_{it,\alpha}\tilde{X}'_{it,\star}]\beta_{\star,o} \\ &= \tilde{X}'_{it,\gamma}R'_{\alpha,\gamma}\left(R_{\alpha,\gamma}E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\gamma}]R'_{\alpha,\gamma}\right)^{-1} R_{\alpha,\gamma}E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\star}]\beta_{\star,o}\ .\end{aligned} \tag{19}$$

Note that $\tilde{Y}^*_{it,\gamma}(\mathcal{W}_i) = Y^*_{it,\alpha}(\mathcal{W}_i)$ is true if

$$R'_{\alpha,\gamma}\left(R_{\alpha,\gamma}E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\gamma}]R'_{\alpha,\gamma}\right)^{-1} R_{\alpha,\gamma} = \left(E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\gamma}]\right)^{-1}, \tag{20}$$

which would hold in general if $R_{\alpha,\gamma}$ were symmetric and invertible. However, by definition it is not a square matrix.

We now consider simple example to illustrate this point. Suppose that $\mathcal{M}_\alpha$ is the annual mean model and $\mathcal{M}_\gamma$ is the quarterly mean model. Then $E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\gamma}]$ is the within variance-covariance matrix of the quarterly means, and $E[\tilde{X}^2_{it,\alpha}]$ is the within-variance of the annual mean, which is a weighted average of the quarterly means. Clearly, the "variability" is not in general the same for the higher- and lower-frequency mean, unless we impose some restrictive assumptions. For instance, if we require that the within-variance is the same for all quarterly means and that there is no within-covariance between the quarterly means, then $E[\tilde{X}_{it,\gamma}\tilde{X}'_{it,\gamma}] = E[\tilde{X}^2_{it,\alpha}]I_{k_\gamma}$, where $E[\tilde{X}^2_{it,\alpha}] > 0$. This would imply that the within-variance of summer and winter average temperatures are the same and that there is no inter-seasonal correlation in temperature. These are unrealistic assumptions that we entertain to illustrate our point. In this example, (20) simplifies to

$$\begin{aligned} R'_{\alpha,\gamma}\left(R_{\alpha,\gamma}R'_{\alpha,\gamma}\right)^{-1} R_{\alpha,\gamma} &= I_{k_\gamma}, \\ \frac{1}{\sum_{j=1}^4 |Q_j|^2/H^2} R'_{\alpha,\gamma}R_{\alpha,\gamma} &= I_{k_\gamma}. \end{aligned} \tag{21}$$

The above equality is trivially fulfilled if $R_{\alpha,\gamma}$ is proportional to the identity matrix, which would imply that both models are identical. But this is not true in this simple example. If we further simplify the problem by assuming that $|Q_j| = H/4$ for $j = 1, \ldots, 4$, then $\mathcal{R}_{\alpha,\gamma} = \frac{1}{4}\mathbf{1}'_k$, where $\mathbf{1}_k$ is a $k \times 1$ vector with all elements equal to one. It follows that the above equality clearly does not hold, since its left-hand side would simplify to $\frac{1}{4}\mathbf{1}_k\mathbf{1}_k'$. Hence, even in this simple example, it is difficult to show that it is possible to obtain identical predictions of the outcome variable given $\mathcal{W}_i$ when considering two models that do not nest $\mathcal{M}_\star$.

We use the above insights to inform our simulation design.

4 Simulation Study

In this section, we first compare the finite-sample performance of the MCCV and GICs in a baseline design that exhibits a high signal-to-noise ratio. Then, we examine the performance of the model selection criteria for different signal-to-noise ratios.

4.1 Baseline Results: Comparing MCCV and Generalized Information Criteria

Here we illustrate the aforementioned theoretical results using a simple simulation study. To simplify illustration, we only present the simulation results for selecting between the annual mean (A), quarterly mean (Q) and quadratic in annual mean ($QinA$) models. We evaluate the behavior of the model selection criteria for selecting among a broader set of models including bi-annual mean, monthly mean, and temperature bin models in the supplementary appendix.

For each of the DGPs we consider, we use a random sample of counties from the NCDC temperature dataset for the years 1968-1972 as $\mathcal{W}_{it}$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $T = 5$. For each simulation replication, we generate $a_i|\mathcal{W}_{i1}, \mathcal{W}_{i2}, \ldots, \mathcal{W}_{i5} \overset{i.i.d.}{\sim} N(0.5\bar{W}_i, 1)$, where $\bar{W}_i = \sum_{t=1}^{T}\sum_{\tau=1}^{H} W_{it\tau}/(TH)$. The idiosyncratic shocks u_{it} are generated as a bivariate mixture normal that is heteroskedastic and serially correlated as follows. Let $u_i = (u_{i1}, \ldots, u_{iT}) = \epsilon_i^1 + \epsilon_i^2$, where $\epsilon_i^1|\mathcal{W}_{i1}, \ldots, \mathcal{W}_{i5}, a_i \overset{i.i.d.}{\sim} N(-0.5, \Sigma_1)$ and $\epsilon_i^2|\mathcal{W}_{i1}, \ldots, \mathcal{W}_{i5}, a_i \overset{i.i.d.}{\sim}$

$N(0.5, \Sigma_2)$, with

$$\Sigma_1 = \begin{pmatrix} 1 & 0.5 & 0.1 & 0 & 0 \\ 0.5 & 1 & 0.5 & 0.1 & 0 \\ 0.1 & 0.5 & 1 & 0.5 & 0.1 \\ 0 & 0.1 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.1 & 0.5 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 1 & 0.5 & 0.1 & 0 & 0 \\ 0.5 & 0.75 & 0.5 & 0.1 & 0 \\ 0.1 & 0.5 & 1 & 0.5 & 0.1 \\ 0 & 0.1 & 0.5 & 0.75 & 0.5 \\ 0 & 0 & 0.1 & 0.5 & 1 \end{pmatrix}. \tag{22}$$

The following response functions generate Y_{it} for the three DGPs we consider:

- Annual Mean (A): $Y_{it} = \bar{W}_{it} + a_{i,\alpha} + u_{it,\alpha}$,

- Quadratic in Annual Mean ($QinA$): $Y_{it} = 0.2\bar{W}_{it} - 0.05\bar{W}_{it}^2 + a_{i,\delta} + u_{it,\delta}$,

- Quarterly Mean (Q): $Y_{it} = -0.25\bar{W}_{it}^{Q_1} + 0.75\bar{W}_{it}^{Q_3} + a_{i,\gamma} + u_{it,\gamma}$.

Given the importance of the pseudo-true parameter values as well as the MSE evaluated at these values in our theoretical analysis, we simulate these quantities for models A, Q, and $QinA$ using 2000 simulation replications using the sample of all counties in our dataset ($n = 3078$) to ensure that our simulated quantities are as close as possible to their population analogues. Table 1 presents the simulation mean of coefficients ($\bar{\beta}_\alpha$) and MSE estimated using $\bar{\beta}_\alpha$ for our entire sample, i.e. $MSE(\bar{\beta}_\alpha) = \sum_{i=1}^n \sum_{t=1}^T (\tilde{y}_{it} - \tilde{x}'_{it,\alpha}\bar{\beta}_\alpha)^2/(nT)$, for all three models we consider when the DGP is A, Q, and $QinA$, respectively. Note that when $QinA$ is the DGP, the annual mean (A) and quarterly mean (Q) models yield very similar MSE at $\bar{\beta}_\alpha$. Similarly, when Q is the DGP, the MSE at $\bar{\beta}_\alpha$ is similar for models A and $QinA$. However, the predicted values of the outcome given the models' pseudo-true parameter values are quite different. Hence, our theoretical results would predict that when selecting between A and Q (A and $QinA$) when the DGP is $QinA$ (Q), we expect BIC to be pseudo-inconsistent, i.e. it will choose the larger model among the two models under consideration.

For each DGP, we examine the performance of the following model selection criteria:

Table 1: Simulation Mean of Model Coefficients and Mean Squared Error

DGP:		A		$QinA$		Q	
$\mathcal{M}_\alpha$	$X^k_{it,\alpha}$	$\bar{\beta}^k_\alpha$	MSE($\bar{\beta}_\alpha$)	$\bar{\beta}^k_\alpha$	MSE($\bar{\beta}_\alpha$)	$\bar{\beta}^k_\alpha$	MSE($\bar{\beta}_\alpha$)
A	A	1.000	0.59	-5.218	1.05	0.112	1.56
$QinA$	A	1.002	0.59	0.202	0.59	0.545	1.56
	A^2	0.000		-0.050		-0.004	
Q	Q1	0.249	0.59	-1.314	1.05	-0.250	0.59
	Q2	0.246		-1.206		0.000	
	Q3	0.248		-1.307		0.750	
	Q4	0.249		-1.305		0.000	

Notes: The table presents $\bar{\beta}^k_\alpha$, the simulation mean for each estimated element of the parameter vector in the models considered across 2000 simulation replications for each DGP (A, $QinA$ and Q). In this design, $n = 3078$ (the total number of counties in the dataset) and $T = 5$. We use $\bar{\beta}^k_\alpha$ to calculate the mean squared error $\text{MSE}(\bar{\beta}_\alpha) = \sum_{i=1}^n \sum_{t=1}^T (\tilde{y}_{it} - \tilde{x}'_{it,\alpha}\bar{\beta}_\alpha)^2/(nT)$.

- MCCV (n_c/n)
  - $n_c/n = p = 0.75$ (MCCV with fixed training to full sample ratios, hereinafter MCCV-p),
  - $n_c = n^{-1/4}$ (MCCV with fixed training to full sample ratios, hereinafter MCCV-Shao);
- $GIC_{\alpha,\lambda_{nT}} = -n(T-1)\log(\hat{\sigma}^2_\alpha) - \lambda_{nT} k_\alpha$, where $\hat{\sigma}^2 = \sum_{i=1}^n \sum_{t=1}^T (\tilde{y}_{it} - \tilde{x}'_{it,\alpha}\hat{\beta}_\alpha)^2/(nT)$,
  - $\lambda_{nT} = 2$ (AIC),
  - $\lambda_{nT} = \log(nT)$ (BIC),
  - $\lambda_{nT} = \sqrt{nT\log(\log(nT))}$ (SW$_1$),
  - $\lambda_{nT} = \sqrt{nT\log(nT)}$ (SW$_2$).

We use the same random sample of n counties from the full NCDC sample of 3,074 counties and use the temperature data for these counties between 1968-72 as our high-frequency regressor $\{\mathcal{W}_i\}_{i=1}^n$ in all the simulation designs. The outcome variable is generated using the DGP in question. All regression models are implemented on the generated data and the six model selection criteria are calculated for each model. The simulation probabilities (proportions) of selecting a particular model for each combination are computed using 500 simulation replications.

Figure 1: Simulation Results for the Model Selection Criteria

Notes: For each DGP (indicated in the first row), the figure plots the simulation probability (proportion) that a particular model (A, Q or $QinA$) is chosen by a model selection criterion in a given model selection problem. The model selection problems we consider are listed in the first column. The model selection criteria are given in the second column. For $n = 500, 3000$, the number of simulation replications is 500.

For each DGP we consider, Figure 1 shows the model selection simulation probability for three different model selection problems when $n = 500, 3000$: (i) A, Q, (ii) $A, QinA$, and (iii) $A, Q, QinA$. When all models are considered as in (iii), AIC and MCCV-p are not model selection consistent as our theoretical results predict; they specifically select overfit models, such as Q or $QinA$, even when $\mathcal{M}_\star$ (corresponding to A) is among the models under consideration. In this setting, MCCV-Shao, BIC, SW_1 and SW_2 select the most parsimonious correctly specified model with simulation probability very close or equal to 1.

To study the pseudo-consistency of the model selection criteria, we examine two designs: (1) choose between A and $QinA$ where DGP= Q, (2) choose between A and Q where DGP= $QinA$. Since in both cases the two models under consideration have similar MSE($\bar{\beta}_\alpha$), A should be chosen in both cases. BIC and MCCV-Shao choose the larger model, $QinA$ in (1) and Q in (2), with probability almost equal to 1 when $n = 3000$, whereas SW_1 and SW_2 choose A. Hence, the former criteria exhibit performance consistent with pseudo-inconsistency in model selection. The BIC's pseudo-inconsistency is predicted by our theoretical results. According to Shao (1997), MCCV-Shao and BIC are asymptotically equivalent, hence it is not surprising that they both behave similarly in the simulations.

4.2 Signal-to-noise Ratio and Model Selection Consistency

Given the potential sensitivity of the finite-sample performance of model selection procedure to the signal-to-noise ratio in the DGP,[7] we examine the behavior of the model selection criteria when we vary the signal-to-noise ratio in our design. To do so, for a given DGP, $\mathcal{M}_\star$ in the form of $Y_{it} = X'_{it,\star}\beta_\star + \alpha_i + v_{it}$. To generate v_{it}, we first generate u_{it} as in our baseline design and obtain $\tilde{X}_{it}$ and $\tilde{u}_{it}$ which are demeaned versions of X_{it} and u_{it}. We then construct the rescaled error term v_{it} such that its demeaned version $\tilde{v}_{it}$ satisfies

$$\tilde{v}_{it} = \sqrt{\frac{1}{\rho} \cdot \frac{Var(\tilde{\mathbf{X}}_{it})}{Var(\tilde{u}_{it})}} \times \tilde{u}_{it},$$

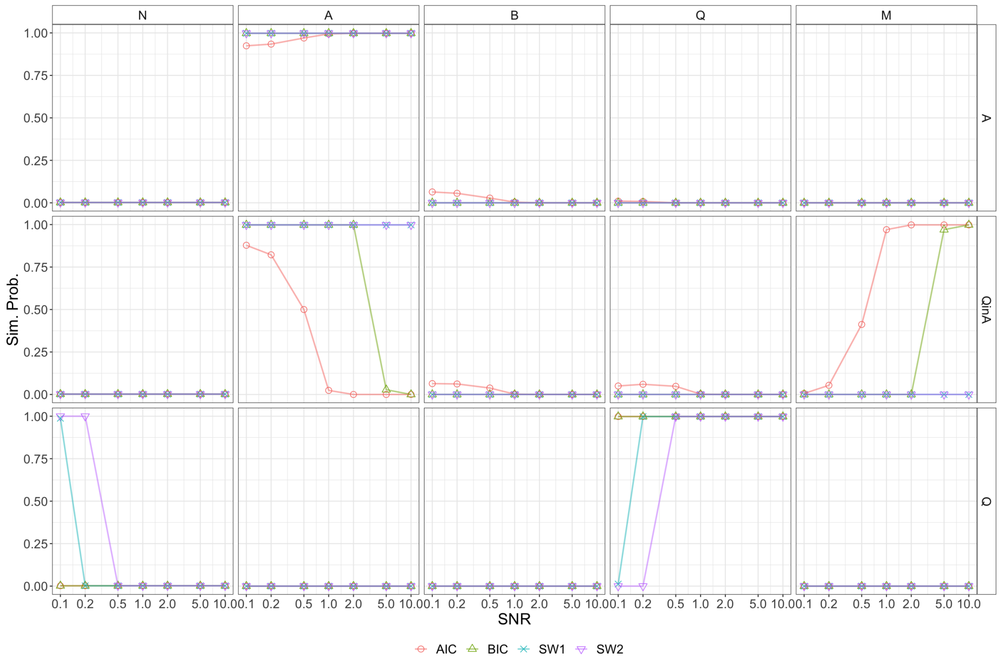
where ρ is the value of SNR we use in a given design.

Figure 2 presents the simulation probability of selecting each model under consideration for two model selection problems from the baseline design: (1) the full set of nested models, (2) the full set of possibly non-nested models. Since we vary the signal-to-noise ratio (SNR) between 0.1 and 10, we also include the null model (with no temperature variable) for the two model selection problems we consider. To simplify illustration, we only include results for AIC, BIC, SW_1 and SW_2, omitting the MCCV procedures.

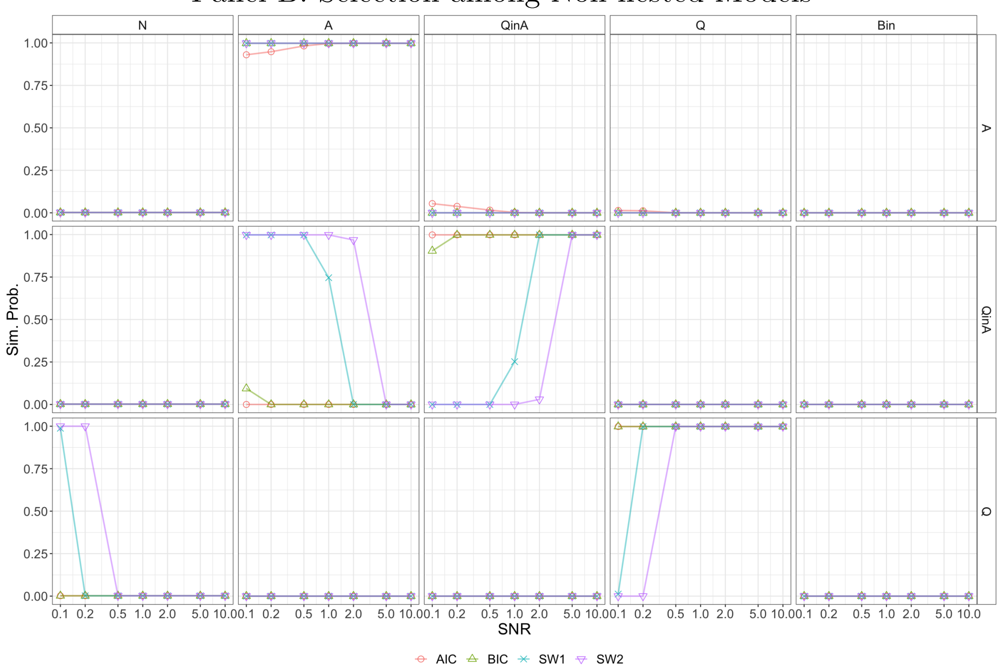
[7] For instance, Hastie et al. (2020) compare the finite-sample performance of best subset selection, LASSO and forward step-wise regression via simulations and illustrate how their relative performance depends on the SNR.

Figure 2: Simulation Results for the Model Selection Criteria under Different SNR

Panel A: Selection among Nested Models



Panel B: Selection among Non-nested Models



Notes: For each DGP (indicated on the right), the figure plots the simulation probability (proportion) that a particular model (indicated on the top) is chosen by a model selection criterion in a given model selection problem under a specific signal-to-noise ratio (SNR). Four information criteria are considered: AIC, BIC, SW1, and SW2. The scale of the horizontal axis is log-transformed.

Overall, the results for high SNRs are consistent with our baseline results. For low SNRs, however, we find that SW_1 and SW_2 can underfit, even if the true DGP is under consideration. For instance, in the second row of Panel B in Figure 2, for SNRs of up to 1.0 or 2.5, the SW criteria tend to select the annual mean model with simulation probability equal or close to one, whereas the AIC and BIC choose the true DGP with high probability. In the third row of Panel B in Figure 2, the SW criteria select the null model with probability close to one for very low SNRs (0.1-0.2).

A practical implication of these simulation results is that the SNR is an important quantity to report when interpreting the results of model selection criteria. In addition, they suggest that rather than reporting the results of a single model selection criterion chosen by the empirical researcher as in current empirical practice, several model selection criteria should be reported to aid the interpretation of their results as we illustrate in our applications.

5 Empirical Applications

The simulation results suggest that the performance of different model selection criteria varies with the relative predominance of signal versus noise in the model. In this section, we provide two empirical applications with starkly different levels of signal-to-noise ratios. The first application revisits the relationship between crop yields and temperature which exhibits a high signal-to-noise ratio, whereas the second application re-examines the GDP-temperature relationship which suffers from a low signal-to-noise ratio.

5.1 Temperature and Crop Yields

The agronomic literature has documented that the accumulation of heat is only beneficial to crop growth over certain ranges of the temperature, and becomes detrimental otherwise (Ritchie and Nesmith 1991). Previous statistical analyses also find evidence of nonlinearity in crop yield response to temperature under different estimation specifications (e.g., Burke and Emerick 2016; Gammans et al. 2017; Schlenker and Roberts 2009). However, the qualitative similarity of nonlinearity does not diminish the importance of exploring the quantitative difference between alternative specifications, especially considering that nuances in the estimation results could be substantially magnified when it comes to projecting future climate impacts.

In this empirical application, we consider different specifications of temperature variables in the following model,

$$\log(Y_{it}) = X'_{it,\alpha}\beta_\alpha + \theta_1 P_{it} + \theta_2 P_{it}^2 + \delta_{1,s}t + \delta_{2,s}t^2 + a_i + \epsilon_{it},$$

where Y_{it} represents corn yields (bushels/acre) in county i in crop year t. The regressors $X_{it,\alpha} = \mu_\alpha(\mathcal{T}_{it})$ represent temperature variables constructed based on daily average temperature of the growing season $\mathcal{T}_{it} \equiv \{T_{it\tau}\}_{\tau=1}^{H}$.[8] P_{it} represents growing-season total precipitation, $\delta_{1,s}$ and $\delta_{2,s}$ characterize state-level quadratic trends, α_i represents county fixed effect and ϵ_{it} is the error term.

We consider the following set of temperature specifications: (a) reference model with no temperature variables, (b) monthly average temperatures, (c) 1°C daily temperature bins, (d) 3°C step function, (e) degree days in the fashion of Schlenker et al. (2006) (SHF degree days, hereafter), (f) piecewise linear function with one knot, and (g) piecewise linear function with two knots. Models (a)-(f) are model candidates considered in Schlenker and Roberts (2009), and model (g) is a more flexible variant of (f). All the models above only consider growing-season temperatures. The last two piecewise linear specifications rely on knots selected by minimizing MSE.[9] Although the specifications considered here are not exhaustive, we believe they are sufficiently rich to illustrate the differentiated performances among different model selection criteria.

We obtain county-level corn yields covering 1950-2015 from USDA Quick Stats. The source of historical weather information is the PRISM dataset, which provides spatially gridded daily data at 4km-by-4km resolution. We follow the data managing procedure in

[8] Previous studies have shown the importance of considering within-season temperature variation in modeling the response of crop yields since the seminal work in Schlenker et al. (2006) and Schlenker and Roberts (2009).

[9] We present the smallest ten MSEs in Appendix Table A1.

Schlenker and Roberts (2009) and obtain county-level daily temperature and precipitation over 1950-2015. Based on the merged county-level data, we first conduct estimation using an unbalanced panel of all available observations covering 1950-2015. The unbalanced sample contains 2,278 counties with a total of 120,995 observations. The estimation results, as reported in Appendix Table A2, are in line with previous findings. We also report signal-to-noise ratios of each estimation model, and these ratios are mostly close to 0.40.[10]

Conducting model selection on the specifications considered, we find that AIC and BIC select the 3°C step function, which is also the preferred specification in Schlenker and Roberts (2009). The two cross-validation procedures and SW_1 select the two-knot piecewise function (knots at 24°C and 26°C), and SW_2 selects the one-knot piecewise linear function (knot at 29°C).

In panel A of Figure 3, we plot the three estimated response functions chosen by different model selection criteria considered. Along with the estimates, we plot 95% confidence intervals constructed by applying the delta method on state-clustered robust standard errors. The general patterns are similar across the three response functions. The 3°C step function is most flexible as it picks up potential nonlinearities in every 3°C. A clear downside of the 3°C step function is its lower precision compared with the piecewise functions, especially for estimates of the high temperatures. Comparing between the two piecewise functions, the two-knot function exhibits a peak positive effect over 24-26°C that is not emphasized in the one-knot function. On high temperature effects, the two piecewise functions deliver similar results but the one-knot function is more precisely estimated. Furthermore, the three models deliver very similar climate change projections as illustrated in Figure 4.

We repeat this empirical application with only using a balanced panel of counties that always planted corn during the sample period. The balanced panel contains 679 counties in the core region of the corn belt, with a total of 44,818 observations. The estimation results, presented in Appendix Table A3, are similar to those obtained using the unbalanced panel, with generally higher signal-to-noise ratios. Our exercise of the model selection indicates

[10]We consider all weather variables as the signal component, and we obtain signal-to-noise ratios by projecting out all the time trends and fixed effects.

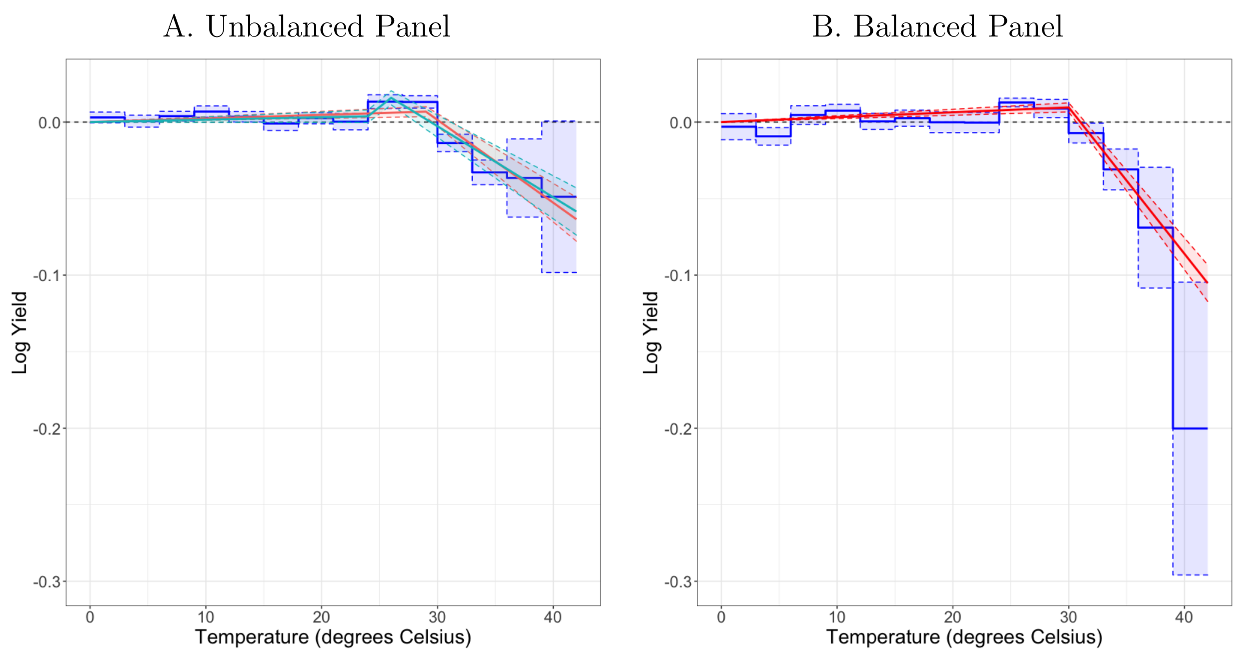


Figure 3: Yield Response to Growing-season Temperature: Selected Models

Notes: The solid lines represent point estimates of the yield response functions, and the shallow bands are 95% confidence intervals constructed by applying the delta method on state-clustered standard errors.

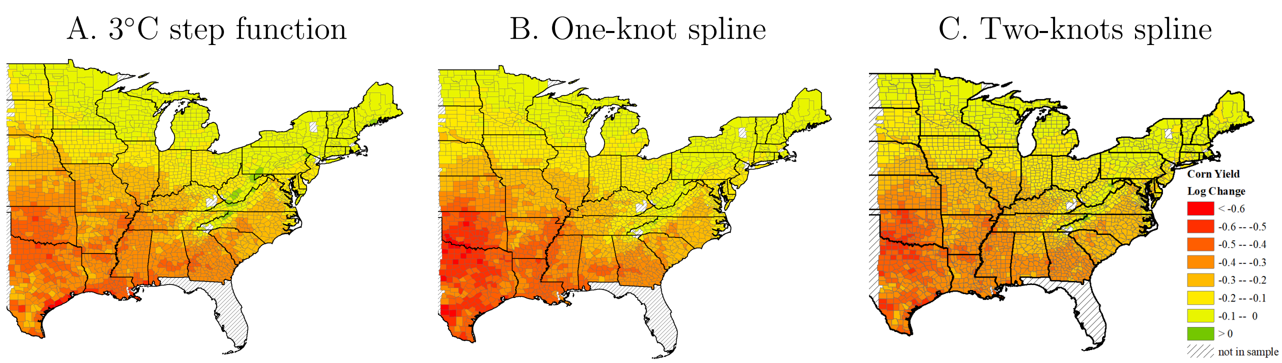


Figure 4: Yield Impacts Projected under Future Climate

Notes: The county-level log changes in corn yields are obtained by applying different yield-response functions to the climate of 2050 projected under HadCM3-B1. The yield-response functions considered here are the three models selected by the different model selection criteria in the unbalanced case.

that both AIC and BIC still select the 3°C step function, while the two SW criteria and the two MCCV procedures all choose the one-knot piecewise function, in which the knot is endogenously determined at 30°C with a minimized MSE. As shown in panel B of Figure 3, although the 3°C step function and the one-knot piecewise function display a similar pattern in the estimated effects, the piecewise function is much more precisely estimated.

Table 2: Model Selection Criteria for the Yield-Temperature Relationship

Model	$\widehat{R^2}$	SNR	AIC	BIC	MCCV$_{\text{-p}}$	MCCV$_{\text{-Shao}}$	SW$_1$	SW$_2$
Unbalanced								
a. no temperature var	9.18%	10.11%	-333091	-332490	0.07126	0.07369	-299391	-259436
b. monthly avg temp	20.19%	25.29%	-348417	-347757	0.06443	0.06752	-311455	-267634
c. 1°C daily temp bin	26.88%	36.76%	-358753	-357792	0.05979	0.06218	-304941	-241143
d. 3°C step function	29.34%	41.52%	-362864	-362127	0.05768	0.05984	-321554	-272577
e. SHF degree days	27.60%	38.13%	-361052	-360421	0.05795	0.05990	-325721	-283833
f. Piecewise: 0-29, 29+	28.99%	40.82%	-361251	-360630	0.05778	0.05954	-326463	-285220
g. Piecewise: 0-24, 24-26, 26+	28.24%	39.35%	-362102	-361471	0.05744	0.05923	-326771	-284883
Balanced								
a. no temperature var	10.25%	11.42%	-138208	-137895	0.04784	0.04998	-126545	-113340
b. monthly avg temp	23.19%	30.19%	-145067	-144702	0.04245	0.04585	-131460	-116054
c. 1°C daily temp bin	36.18%	56.69%	-153185	-152558	0.03530	0.03797	-129858	-103447
d. 3°C step function	39.27%	64.66%	-155416	-154980	0.03322	0.03549	-139217	-120876
e. SHF degree days	38.25%	61.94%	-153302	-152962	0.03583	0.03787	-140667	-126361
f. Piecewise: 0-30, 30+	38.37%	62.25%	-154706	-154375	0.03310	0.03486	-142395	-128455
g. Piecewise: 0-29, 29-33, 33+	36.26%	56.88%	-154789	-154449	0.03315	0.03500	-142153	-127848

Notes: $\widehat{R^2}$ and SNR are calculated based on the following regression: $\widehat{\log(y_{it})} = \widehat{X'_{it,\alpha}}\beta_\alpha + \theta_1 \widehat{P_{it}} + \theta_2 \widehat{P^2_{it}} + \widehat{\epsilon_{it}}$, where the hatted variables are obtained by projecting the original variables out of county and year fixed effects and state quadratic trends. The SNR is formed by dividing the model sum of squares by the residual sum of squares. The columns for MCCV-p and MCCV-Shao presents MSE for cross-validation with 1,000 simulations.

Estimating temperature effects on crop yields corresponds to a case with relatively high predominance of signal in the empirical literature on climate change impact estimation. This statistical high signal is supported by the agronomic knowledge that the varying heat over a growing season constitutes a critical input for crop growth. As we show through simulations, the SW criteria more likely deliver consistent model selection results when the signal-to-noise ratio is relatively high. Reflected in our crop yield application, the SW criteria select the more parsimonious piece-wise linear functions. From an empirical perspective, they outperform the step functions chosen by AIC and BIC, since the piece-wise linear functions have successfully characterized the well-established nonliearity in temperature response with relatively higher precision.

5.2 Temperature and GDP

A different context that features relatively low signal-to-noise ratios is to explain temperature impacts on growth with country-level macro panel data. Using data of 125 countries over 1950-2005, Dell et al. (2012) document that rising temperatures reduce economic growth in poor countries through influencing agricultural and industrial outputs as well as political stability, but rich countries are found to be more immune to higher temperatures. In contrast, using similar data, Burke et al. (2015) (BHM, hereafter) show that the temperature-GDP relationship follows a concave quadratic function, and they argue that the temperature effect on economic production is nonlinear globally. Results in BHM imply that both rich and poor countries will experience substantial economic losses under future climate. The model uncertainty embedded in these analyses would directly influence the policy prescriptions for governments to deal with climate change (Newell et al. 2021).

In this section, we formally evaluate the model selection problem using BHM's original data. The source data contain country-level per capita GDP over 1960-2012 from the World Bank's World Development Indicators, and annual temperature and precipitation over 1900-2010 aggregated to the country level from 0.5 degree gridded monthly data from the University of Delaware reconstruction. With missing values, the merged country-level dataset is an unbalanced panel of 166 countries that covers 1960-2010.

Focusing on the response function of temperature, we maintain BHM's regression specification of terms other than temperature. The panel fixed effect estimation is expressed as follows.

$$\log(Y_{it}) = h(T_{it}) + \theta_1 P_{it} + \theta_2 P_{it}^2 + \delta_{1,i} t + \delta_{2,i} t^2 + a_i + \mu_t + \epsilon_{it}. \tag{23}$$

Y_{it} represents per capita GDP in country i in year t. $h(T_{it})$ denotes an unknown function of annual temperature T_{it}. The role of precipitation is characterized as a quadratic function of annual precipitation P_{it}. Country-level smooth trends are captured by $\delta_{1,i} t + \delta_{2,i} t^2$, and a_i

and μ_t are country and year fixed effects, respectively. ϵ_{it} is the error term.

We consider the following set of specifications of annual temperature: (a) no temperature variable at all, (b) simple average, (c) quadratic, (d) cubic, (e) fourth order, (f) linear spline with knots at every 5°C from 0°C to 25°C, (g) linear spline with knots at every 3°C from 0°C to 27°C. (h) linear spline with one knot determined by minimizing MSE. These models follow the model candidates of higher-order polynomials and multi-knot splines in the robustness evaluation in Burke et al. (2015).[11] We then evaluate the performance of different models with the four GICs (AIC, BIC, SW_1, SW_2) and the two cross-validation procedures (MCCV-p and MCCV-Shao, both with 1,000 simulations).

Following the practices in BHM, we first conduct the regressions above using the unbalanced panel that contains 166 countries with a total of 6,584 country-year paired observations. In the one-knot spline specification, a knot at 16°C is selected as it delivers the minimum MSE (see Appendix Table A5). For all the specifications, we report estimates with country-clustered robust standard errors in Appendix Table A4. To illustrate the results, we plot the estimated temperature functions in Figure 5 using the point estimates obtained from the regressions. Over the support of observed annual temperatures, the quadratic, cubic, and splines specifications yield qualitatively similar results: economic growth increases with temperature in cool areas but decreases with temperature in hot areas. The fourth-order model overshoots when temperature becomes higher than 15°C, and the simple average model suggests that temperature effect is close to zero.

Next, we examine the model selection criteria in Table 3. For the full (unbalanced) sample, AIC selects the most flexible 3°C segment spline model. BIC and the two MCCV procedures all select the one-knot linear spline model. The estimates of this specification suggest that a 1°C warming below 16°C is associated with a 0.63% increase in GDP, and a 1°C warming above 16°C is associated with a 1.16% reduction in GDP. These two models exhibit very similar response functions, see the left panel in Figure 5. When turning to the SW criteria, we find that they both select the reference model with no temperature variables.

[11] We use linear splines instead of restricted cubic splines due to degree-of-freedom concerns, and we adapt the spline model to endogenously selecting the knot in model (h).

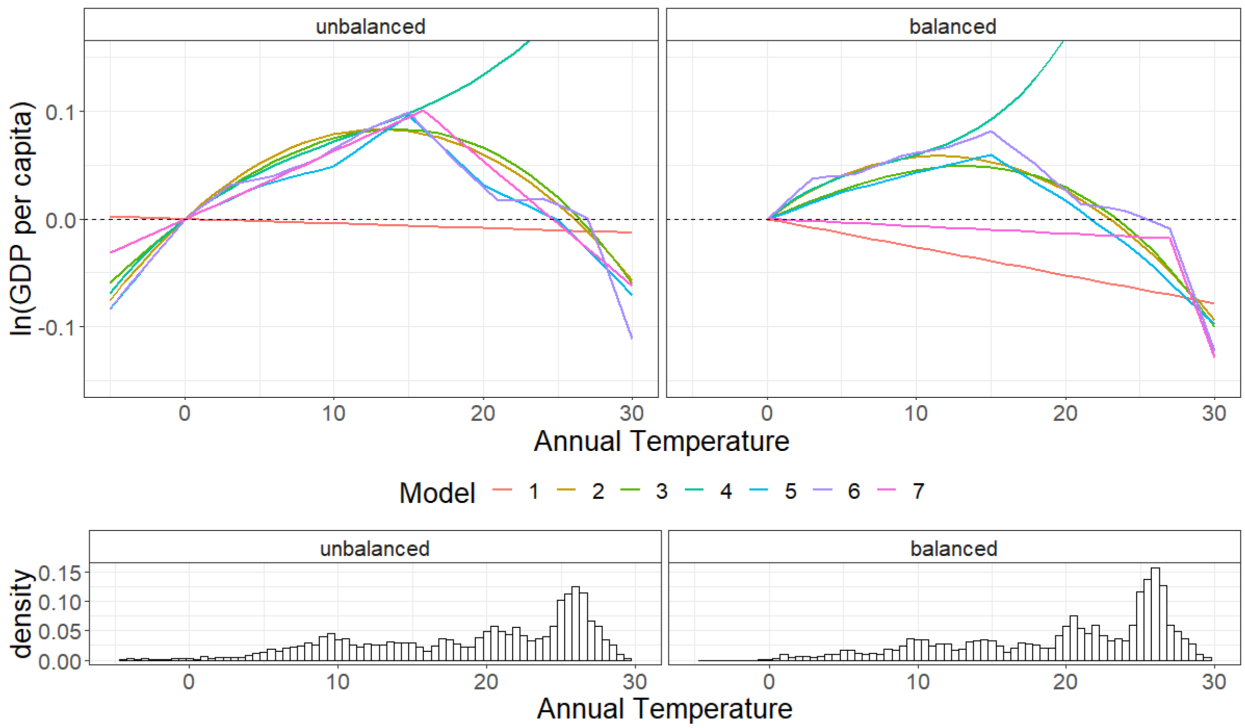


Figure 5: Estimated Functions of Temperature-GDP Relationship: Point Estimates

Notes: The upper panels plot point estimates of the temperature-GDP relationship obtained from the unbalanced and balanced samples, respectively. The lower panels plot the distributions of annual temperatures in the unbalanced and balanced samples, respectively. Models 1-7 in the figure are (1) simple average, (2) quadratic, (3) cubic, (4) fourth-order, (5) 5°C segment spline, (6) 3°C segment spline, and (7) one-knot spline with the knot selected by minimizing MSE.

For the balanced sample, the models selected by the different criteria reflect some of the patterns in our simulation results. Our simulations have shown that the SW criteria tend to underfit models when signal-to-noise ratios are low. Indeed, as presented in Table 3, all the models for estimating GDP feature signal-to-noise ratios below 1%, far below those in the crop yields application. In this case, our simulations suggest that the models selected by other information criteria like BIC may better approximate the underlying relationship. The nonlinear relationship characterized by the one-knot piecewise function (or the 3°C step function) also qualitatively resembles the quadratic function favored by BHM.

A. Unbalanced B. Balanced

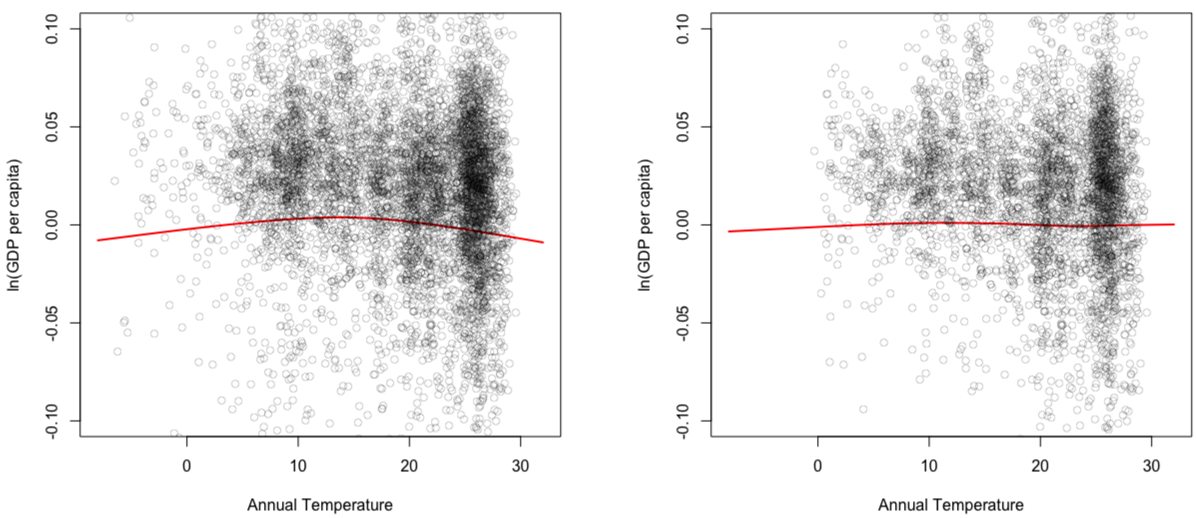


Figure 6: Estimated Smoothing Splines: Temperature and GDP

Notes: The red curves are fitted smoothing splines, and the round circles are overlaid observations. For illustration purpose, we restrict the presentation of the observations to those with ln(GDP per captia) in between -0.10 and 0.10.

Since many countries have a substantial amount of missing values, we examine the results for the balanced panel of 86 countries with a total of 4300 observations. The estimation results are reported in Appendix Table A6, and the temperature functions recovered from the point estimates are plotted in the right panel of Figure 5. Because the balanced sample only contains very few observations with temperature below 0°C, we suppress the spline segment below 0°C in estimating the 5°C and 3°C segment splines. Most of the results for the balanced sample are qualitatively similar to those obtained with the unbalanced sample. However, an estimated pattern that sharply contrasts with the unbalanced results is the one-knot spline function. When evaluated based on the balanced sample, the knot minimizes MSE becomes 27°C, and the estimated temperature effect is close to zero before 27°C and drops substantially beyond 27°C.[12] This is the model that is selected by AIC. BIC selects the quadratic model (the preferred specification in BHM). The two MCCV procedures select the most flexible 3°C segment spline model. Unlike the results using the unbalanced sample, the models selected by AIC, BIC and the MCCV procedures do not deliver qualitatively similar response functions, see the right panel of Figure 5. Specifically, the one-knot spline function suggests that only temperature increases above 27°C adversely affect GDP, while the 3°C segment spline and the quadratic model yield an inverted U-shaped response function. As for the SW criteria, both select the reference model with no temperature variables.

Since the functional forms considered are not motivated by economic theory, we further turn to a fully data-driven approach to detect if it delivers estimated patterns that resemble certain functional forms we examined above. Specifically, we model $h(T_{it})$ in equation (23) as a smoothing spline.[13] Figure 6 shows the estimated splines based on the unbalanced and the balanced samples, respectively. In the unbalanced case, the spline is similar to a quadratic

[12] We note that for both the unbalanced and the balanced cases, when plotting the spline function over the support of observed annual temperatures, it shows that a large portion of the 95% confidence band overlaps with the zero line (see Appendix Figure A1). These 95% confidence intervals are obtained by applying the delta method using country-clustered robust standard errors.

[13] We set up the estimation as a general additive model (GAM) using R, and implement the estimation using a cubic spline basis given their favorable theoretical properties (Wood 2017).

Table 3: Model Selection Criteria in the Temperature-GDP Relationship

Model	$\widehat{R^2}$	SNR	AIC	BIC	MCCV$_{\text{-p}}$	MCCV$_{\text{-Shao}}$	SW$_1$	SW$_2$
Unbalanced sample:								
a. none	0.09%	0.09%	-37166	-34558	0.00542	0.00690	*8007*	*54457*
b. simple avg	0.09%	0.09%	-37164	-34549	0.00543	0.00690	8126	54698
c. quadratic	0.46%	0.46%	-37186	-34564	0.00526	0.00674	8222	54915
d. cubic	0.47%	0.47%	-37184	-34556	0.00527	0.00676	8341	55155
e. 4th order	0.47%	0.48%	-37183	-34547	0.00529	0.00679	8460	55395
f. 5°C segment spline	0.58%	0.58%	-37183	-34527	0.00519	0.00668	8813	56110
g. 3°C segment spline	0.84%	0.84%	*-37192*	-34509	0.00515	0.00668	9275	57056
h. one-knot spline	0.52%	0.52%	-37189	*-34568*	*0.00511*	*0.00659*	8218	54911
Balanced sample:								
a. none	0.07%	0.07%	-25160	-23747	0.00424	0.00516	*-4387*	*16503*
b. simple avg	0.11%	0.11%	-25160	-23740	0.00423	0.00513	-4293	16691
c. quadratic	0.41%	0.41%	-25172	*-23753*	0.00416	0.00507	-4306	16678
d. cubic	0.42%	0.43%	-25169	-23736	0.00416	0.00508	-4115	17057
e. 4th order	0.44%	0.44%	-25167	-23729	0.00417	0.00510	-4020	17246
f. 5°C segment spline	0.43%	0.43%	-25163	-23711	0.00408	0.00501	-3828	17626
g. 3°C segment spline	0.77%	0.77%	-25169	-23692	*0.00402*	*0.00499*	-3461	18370
h. one-knot spline	0.54%	0.54%	*-25176*	-23750	0.00424	0.00516	-4215	16863

Notes: $\widehat{R^2}$ and SNR are calculated based on the following regression: $\widehat{\log(y_{it})} = \widehat{h(T_{it})} + \theta_1 \widehat{P_{it}} + \theta_2 \widehat{P_{it}^2} + \widehat{\epsilon_{it}}$, where the hatted variables are obtained by projecting the original variables out of country and year fixed effects and country quadratic trends. The SNR is formed by dividing the model sum of squares by the residual sum of squares. The columns for MCCV-p and MCCV-Shao presents MSE for cross-validation with 1,000 simulations. We italicize the smallest model selection criterion in each column for the unbalanced and balanced sample, respectively.

function while the magnitude of the estimates is substantially smaller than that presented in Figure 5. In the balanced case, the spline is very close to the zero line, despite modest curvature around 10°C.

Overall, this application illustrates that, unlike the crop yield example, the GDP-temperature relationship is sensitive to the sample used for estimation as well as the model selection criterion in question. The sensitivity of the results is not surprising when considering the magnitude of the signal-to-noise ratio in this application.

6 Conclusion

This paper formalizes the model selection problem faced by applied researchers and policy-makers interested in examining the climate change impact on outcomes of interest. Conditions for the consistency of Monte Carlo Cross-Validation and generalized information criteria are derived. An interesting takeaway from this analysis is that, even though all models under consideration are linear in the parameters, the model selection problem in this empirical literature is a nonlinear model selection problem. The nonlinearity stems from the fact that all models considered rely on different summary statistics of the underlying high-frequency regressor. As a result, BIC can be pseudo-inconsistent in this context similar to nonlinear model selection problems (Hong and Preston 2012; Sin and White 1996). Simulation analysis of the finite-sample behavior of the model selection criteria points to the importance of the signal-to-noise ratio. Two empirical applications with starkly different signal-to-noise ratios illustrate the practical recommendations implied by our results.

This paper is concerned with the behavior of model selection criteria in the context of climate change impact studies. While it constitutes a first step toward principled model selection in this important empirical context, there are several interesting directions for future research. More flexible procedures to estimate the response functions would be a good substitute to the model selection approach taken in this literature. Allowing for possible nonlinearities between regressors and fixed effects is another important departure from the setup in this paper. Finally, providing valid post-selection inference for the aforementioned methods constitutes a priority for future work.

A Derivations and Proofs

A.1 MSPE Derivation

Let $\tilde{\mathbb{Z}}$ denote "future" values of $\widetilde{\mathbb{Y}}$ whereas $\hat{\beta}(\alpha_\alpha)$ was estimated using the sample $\widetilde{\mathbb{Y}}$ and $\widetilde{\mathbb{X}}_{n,\alpha}$).

$$\begin{aligned}\widehat{\Gamma}_{\alpha,nT} &= \frac{1}{nT}\left\|\widetilde{\mathbb{Z}} - \widetilde{\mathbb{X}}_\alpha \hat{\beta}_\alpha\right\|^2 \\ &= \frac{1}{nT}\left\|\widetilde{\mathbb{U}}_z + \widetilde{\mathbb{X}}_\star \beta_{\star,o} - \underbrace{\widetilde{\mathbb{X}}_\alpha \left(\widetilde{\mathbb{X}}_\alpha' \widetilde{\mathbb{X}}_\alpha\right)^{-1} \widetilde{\mathbb{X}}_\alpha'}_{\equiv P_\alpha} \widetilde{\mathbb{Y}}\right\|^2 \\ &= \frac{1}{nT}\left\|\widetilde{\mathbb{U}}_z + (I_{nT} - P_\alpha)\widetilde{\mathbb{X}}_\star \beta_{\star,o} - P_\alpha \widetilde{\mathbb{U}}\right\|^2 \end{aligned} \tag{24}$$

where $\widetilde{\mathbb{U}}_z$ denotes the within-individual demeaned error term of the observations $\widetilde{\mathbb{Z}}$.
Let $\Gamma_{\alpha,nT}$ denote the expectation of $\widehat{\Gamma}_{\alpha,nT}$ conditional on $\{\mathcal{W}_i\}_{i=1}^n$

$$\Gamma_{\alpha,nT} = \frac{1}{nT} E[\widetilde{\mathbb{U}}_z' \widetilde{\mathbb{U}}_z | \{\mathcal{W}_i\}_{i=1}^n] + E[\widetilde{\mathbb{U}}' P_\alpha \widetilde{\mathbb{U}} | \{\mathcal{W}_i\}_{i=1}^n] + \frac{1}{nT} \beta_{\star,o}' \widetilde{\mathbb{X}}_\star' (I_{nT} - P_\alpha) \widetilde{\mathbb{X}}_\star \beta_{\star,o} \tag{25}$$

The first term on the right hand size of the equality equals $E\left[\sum_{i=1}^n \sum_{t=1}^T \tilde{u}_{it}^2\right]/nT = E[\tilde{u}_{it}^2] = \sigma^2(T-1)/T$. The second term can be simplified as follows

$$\begin{aligned}&\frac{1}{nT} E\left[\widetilde{\mathbb{U}}' \mathbb{X}_\alpha (\widetilde{\mathbb{X}}_\alpha' \widetilde{\mathbb{X}}_\alpha)^{-1} \widetilde{\mathbb{X}}_\alpha' \widetilde{\mathbb{U}} | \{\mathcal{W}_i\}_{i=1}^n\right] = \frac{1}{nT} tr\left(E\left[\widetilde{\mathbb{U}}\widetilde{\mathbb{U}}' \mathbb{X}_\alpha (\widetilde{\mathbb{X}}_\alpha' \widetilde{\mathbb{X}}_\alpha)^{-1} \widetilde{\mathbb{X}}_\alpha' | \{\mathcal{W}_i\}_{i=1}^n\right]\right) \\ =&\frac{1}{nT}\sigma^2 tr(\underbrace{(I_n \otimes (I_T - \mathcal{J}_T/T))}_{\equiv I_n \otimes Q_T} \widetilde{\mathbb{X}}_\alpha (\widetilde{\mathbb{X}}_\alpha' \widetilde{\mathbb{X}}_\alpha)^{-1} \widetilde{\mathbb{X}}_\alpha') = \frac{1}{nT}\sigma^2 k_\alpha. \end{aligned} \tag{26}$$

where the last equality follows by noting $(I_n \otimes Q_T)\widetilde{\mathbb{X}}_\alpha = \widetilde{\mathbb{X}}_\alpha$ as well as properties of the trace.

A.2 Proof of Proposition 1

Proof. The proof is adapted from Shao (1993) to the setting of a fixed effects model with stochastic high-frequency regressors. Following Shao (1993), we first show the results for Balanced Incomplete Cross-Validation (BICV) with stochastic regressors, then we extend the results to MCCV. Let $\mathcal{B}$ be a collection of b subsets of $\{1, \ldots, n\}$ that have size n_v such that (i) for each i, $1 \le i \le n$, the same number of subsets of $\mathcal{B}$ include it, (ii) for each pair (i,j) for $i, j \in \{1, \ldots, n\}$, the same number of subsets of $\mathcal{B}$ include it. From (3.1) in Shao

(1993) and the balance property of $\mathcal{B}$, From (3.1) in Shao (1993) and the balance property of $\mathcal{B}$,

$$\hat{\Gamma}^{BICV}_{\alpha,nT} \geq \frac{1}{n_v T b}\sum_{s\in\mathcal{B}} \|\widetilde{\mathbb{Y}}_s - \widetilde{\mathbb{X}}_{s,\alpha}\hat{\beta}_\alpha\|^2 = n^{-1}\|\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_\alpha\hat{\beta}_\alpha\|^2 = (nT)^{-1}\widetilde{\mathbb{U}}'\widetilde{\mathbb{U}} + \Delta_{\alpha,nT} + o_p(1) \quad (27)$$

where the last equality follows from the proof of (3.5) in Shao (1993). (i) in this proposition follows by letting $R_n = \hat{\Gamma}^{BICV}_{\alpha,nT} - \|\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_\alpha\hat{\beta}_\alpha\|^2/n$.
By Condition 1.3.(iii) with $s \in \mathcal{B}$ in lieu of $s \in \mathcal{R}$, it follows for every $s \in \mathcal{B}$,

$$\begin{aligned}
&\frac{1}{n}\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha - \frac{1}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s} = \frac{1}{n}\left[\widetilde{\mathbb{X}}'_{\alpha,s^c}\widetilde{\mathbb{X}}_{\alpha,s^c} + \widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s} - \frac{n}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right] \\
=&\frac{1}{n}\left[\widetilde{\mathbb{X}}'_{\alpha,s^c}\widetilde{\mathbb{X}}_{\alpha,s^c} + \widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s} - \frac{n_c+n_v}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right] = \frac{1}{n}\left[\widetilde{\mathbb{X}}'_{\alpha,s^c}\widetilde{\mathbb{X}}_{\alpha,s^c} - \frac{n_c}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right] \\
=&\frac{n_c}{n}\left[\frac{1}{n_c}\widetilde{\mathbb{X}}'_{\alpha,s^c}\widetilde{\mathbb{X}}_{\alpha,s^c} - \frac{1}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right] = o_p\left(\frac{n_c}{n}\right) \qquad (28)
\end{aligned}$$

With some further manipulations,

$$\begin{aligned}
\left(\frac{1}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right)^{-1}\frac{1}{n}\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha - I &= o_p\left(\frac{n_c}{n}\right)\left(\frac{1}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right)^{-1} \\
\left(\frac{1}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right)^{-1} - \left(\frac{1}{n}\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha\right)^{-1} &= o_p\left(\frac{n_c}{n}\right)\left(\frac{1}{n_v}\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right)^{-1}\left(\frac{1}{n}\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha\right)^{-1} \\
\left(\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right)^{-1} - \frac{n}{n_v}\left(\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha\right)^{-1} &= o_p\left(\frac{n_c}{n}\right)\left(\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s}\right)^{-1}\underbrace{\left(\frac{1}{n}\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha\right)^{-1}}_{=O_p(1) \text{ by } (3.3)} \qquad (29)
\end{aligned}$$

Hence, together with Condition 1.3.i, the above implies that

$$(\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s})^{-1} - \frac{n}{n_v}(\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha)^{-1} = o_p\left(\frac{n_c}{n}\right)(\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s})^{-1} \qquad (30)$$

For $P_{\alpha,s} = \widetilde{\mathbb{X}}_{\alpha,s}(\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s})^{-1}\widetilde{\mathbb{X}}'_{\alpha,s}$,

$$\begin{aligned}
P_{\alpha,s} =&\frac{n}{n_v}\widetilde{\mathbb{X}}_{\alpha,s}\left(\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha\right)^{-1}\widetilde{\mathbb{X}}'_{\alpha,s} + \frac{n}{n_v}\widetilde{\mathbb{X}}_{\alpha,s}\left(\widetilde{\mathbb{X}}'_\alpha\widetilde{\mathbb{X}}_\alpha\right)^{-1}\widetilde{\mathbb{X}}'_{\alpha,s} + o\left(\frac{n_c}{n}\right)\widetilde{\mathbb{X}}_{\alpha,s}(\widetilde{\mathbb{X}}'_{\alpha,s}\widetilde{\mathbb{X}}_{\alpha,s})^{-1}\widetilde{\mathbb{X}}'_{\alpha,s} \\
=&\frac{n}{n_v}Q_{\alpha,s} + o_p\left(\frac{n_c}{n}\right)P_{\alpha,s} \qquad (31)
\end{aligned}$$

Given that $n_v/n = O(1)$, it follows that

$$Q_{\alpha,s} = P_{\alpha,s}\left(\frac{n_v}{n} + o_p\left(\frac{n_c}{n}\right)\right) \qquad (32)$$

From the balance property of $\mathcal{B}$,

$$\frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\mathbf{r}'_{\alpha,s}Q_{\alpha,s}\nabla_{\alpha,s} = \frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\sum_{i\in s}\sum_{t=1}^{T} w_{it,\alpha}r^2_{it,\alpha} = \frac{1}{n_v Tb}\left(\frac{n_v b}{n} - \frac{n_v b}{n}\frac{n_v-1}{n-1}\right)\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it,\alpha}r^2_{it,\alpha}$$

$$=\frac{1}{T}\left(\frac{1}{n} - \frac{n_v-1}{n(n-1)}\right)\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it,\alpha}r^2_{it,\alpha}$$

where $\mathbf{r}_{\alpha,s} = \widetilde{\mathbb{Y}}_s - \widetilde{\mathbb{X}}_{\alpha,s}\hat{\beta}_\alpha$ and $r_{it,\alpha} = \tilde{Y}_{it} - \tilde{X}'_{it,\alpha}\hat{\beta}_\alpha$.
By (32) and $n_v/n \to 1$ and $n_c \to \infty$, let $c_n = n_v(n+n_c)n_c^{-2}$,

$$\begin{aligned}
\frac{c_n}{n_v Tb}\|P_{\alpha,s}\mathbf{r}_{\alpha,s}\|^2 &= \left[\frac{n_v}{n} + o_p\left(\frac{n_c}{n}\right)\right]^{-1}\frac{c_n}{n_v Tb}\sum_{s\in\mathcal{B}}\mathbf{r}'_{\alpha,s}Q_{\alpha,s}\mathbf{r}_{\alpha,s}\\
&= \left[\frac{n}{n_v} + o_p\left(\frac{n_c}{n}\right)\right]\frac{n_v(n+n_c)n_c^{-2}}{n_v Tb}\sum_{s\in\mathcal{B}}\mathbf{r}'_{\alpha,s}Q_{\alpha,s}\mathbf{r}_{\alpha,s}\\
&= \left[1 + o_p\left(\frac{n_c}{n}\right)\right]\frac{n+n_c}{n_c(n-1)T}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it,\alpha}r^2_{it,\alpha} \qquad (33)
\end{aligned}$$

Now we can write $\hat{\Gamma}^{BICV}_{\alpha,n} = A_\alpha + B_\alpha$, where

$$\hat{\Gamma}^{BICV}_{\alpha,n} = \frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\|(I_{n_vT} - Q_{\alpha,s})^{-1}(Y_s - \widetilde{\mathbb{X}}_{\alpha,s}\hat{\beta}_\alpha)\|^2 = \frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\mathbf{r}'_{\alpha,s}(I_{n_vT} - Q_{\alpha,s})^{-2}\mathbf{r}_{\alpha,s}$$

$$=\frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\|(I_{n_vT} - Q_{\alpha,s})^{-1}(Y_s - \widetilde{\mathbb{X}}_{\alpha,s}\hat{\beta}_\alpha)\|^2$$

$$=\underbrace{\frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\mathbf{r}'_{\alpha,s}(I_{n_vT} - Q_{\alpha,s})^{-1}U_{\alpha,s}(I_{n_vT} - Q_{\alpha,s})^{-1}\mathbf{r}_{\alpha,s}}_{\equiv A_\alpha}$$

$$+\underbrace{\frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\mathbf{r}'_{\alpha,s}(I_{n_vT} - Q_{\alpha,s})^{-1}(I_{n_vT} - U_{\alpha,s})(I_{n_vT} - Q_{\alpha,s})^{-1}\mathbf{r}_{\alpha,s}}_{\equiv B_\alpha}$$

where

$$Z_{\alpha,s} = (I_{n_vT} - Q_{\alpha,s})(I + c_n P_{\alpha,s})(I_{n_vT} - Q_{\alpha,s})$$

From the balance property of $\mathcal{B}$ and (33)

$$\begin{aligned}
A_\alpha &= \frac{1}{n_v Tb}\sum_{s\in\mathcal{B}}\|\mathbf{r}_{\alpha,s}\|^2 + \frac{c_n}{n_v Tb}\sum_{s\in\mathcal{B}}P_{\alpha,s}\|\mathbf{r}_{\alpha,s}\|^2\\
&= \frac{1}{nT}\|\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_\alpha\hat{\beta}_\alpha\|^2 + \left[1 + o_p\left(\frac{n_c}{n}\right)\right]\frac{n+n_c}{n_c(n-1)T}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it,\alpha}r^2_{it,\alpha} \qquad (34)
\end{aligned}$$

Assume $\mathcal{M}_\alpha$ is in Category II. Then by (34) and $\sum_{i=1}^n \sum_{t=1}^T w_{it,\alpha} r_{it,\alpha}^2 = k_\alpha \sigma^2 + o_p(1)$

$$\begin{aligned}
A_\alpha &= \frac{1}{n}\widetilde{\mathbb{U}}'(I-P_\alpha)\widetilde{\mathbb{U}} + \left[1+o_p\left(\frac{n_c}{n}\right)\right] \frac{n+n_c}{n_c(n-1)T}\left[k_\alpha\sigma^2 + o_p(1)\right] \\
&= \frac{1}{n}\widetilde{\mathbb{U}}'\widetilde{\mathbb{U}} - \frac{1}{n}\widetilde{\mathbb{U}}'P_\alpha\widetilde{\mathbb{U}} + \left[1+o_p\left(\frac{n_c}{n}\right)\right] \frac{n+n_c}{n_c(n-1)T}\left[k_\alpha\sigma^2 + o_p(1)\right] \\
&= \frac{1}{n}\widetilde{\mathbb{U}}'\widetilde{\mathbb{U}} + \frac{k_\alpha\sigma^2}{n_cT} + o_p\left(\frac{1}{n_c}\right)
\end{aligned}$$

It remains to show that $B_\alpha = o_p(n_c^{-1})$. From (32)

$$\begin{aligned}
(I_{n_vT}-Q_{\alpha,s})P_{\alpha,s}(I_{n_vT}-Q_{\alpha,s}) &= \left(1-\frac{n_v}{n}+o\left(\frac{n_c}{n}\right)\right)P_{\alpha,s}(I_{n_vT}-Q_{\alpha,s}) = \left(1-\frac{n_v}{n}+o\left(\frac{n_c}{n}\right)\right)^2 P_{\alpha,s} \\
&= \left(\frac{n_c}{n}+o_p\left(\frac{n_c}{n}\right)\right)^2 P_{\alpha,s} \qquad (35)
\end{aligned}$$

Thus,

$$\left(\frac{n}{n_c}\right)^2 (I_{n_vT}-Q_{\alpha,s})P_{\alpha,s}(I_{n_vT}-Q_{\alpha,s}) = (1+o(1))^2 P_{\alpha,s} \geq \frac{1}{2}P_{\alpha,s} \qquad (36)$$

for $s \in \mathcal{B}$ and n sufficiently large. Pre- and post-multiplying the above by $(I_{n_vT}-Q_{\alpha,s})^{-1}$ yields

$$(I_{n_vT}-Q_{\alpha,s})^{-1}P_{\alpha,s}(I_{n_vT}-Q_{\alpha,s})^{-1} \leq 2\left(\frac{n}{n_c}\right)^2 P_{\alpha,s} \qquad (37)$$

Similarly by (32)

$$\begin{aligned}
Z_{\alpha,s} &= \left\{I_{n_vT} - \left[\frac{n_v}{n}+o_p\left(\frac{n_c}{n}\right)\right]P_{\alpha,s}\right\}(I_{n_vT}+c_nP_{\alpha,s})\left\{I_{n_vT} - \left[\frac{n_v}{n}+o_p\left(\frac{n_c}{n}\right)\right]\right\} \\
&= I_{n_vT} + \left[o_p\left(\frac{n_c}{n}\right)\right]^2 (1+c_n)P_{\alpha,s} \qquad (38)
\end{aligned}$$

since $c_n(1-n_v/n)^2 = (2-n_v/n)n_v/n$. Using (34)

$$\begin{aligned}
&(I_{n_vT}-Q_{\alpha,s})^{-1}(I_{n_vT}-Z_{\alpha,s})(I_{n_vT}-Q_{\alpha,s})^{-1} \\
&= \left[o_p\left(\frac{n_c}{n}\right)\right]^2 (1+c_n)(I_{n_vT}-Q_{\alpha,s})^{-1}P_{\alpha,s}(I_{n_vT}-Q_{\alpha,s})^{-1} \leq o_p(1)(1+c_n)P_{\alpha,s}.
\end{aligned}$$

Thus,

$$B_\alpha \leq o_p(1)(1+c_n)\left(\frac{1}{n_vTb}\sum_{s\in\mathcal{B}}\|P_{\alpha,s}\mathbf{r}_{\alpha,s}\|^2\right) = o_p\left(\frac{1}{n_c}\right) \qquad (39)$$

since from the above $(c_n/n_vTb)\sum_{s\in\mathcal{B}}\|P_{\alpha,s}\mathbf{r}_{\alpha,s}\|^2 = O_p(n_c^{-1})$, which proves (ii) in the proposition for BICV. (iii) follows in a straightforward manner from (i) and (ii).

The extension of the proof to MCCV is straightforward from Theorem 2 in Shao (1993) assuming the sufficient conditions given in Condition 1. $\square$

References

J. M. Addoum, D. T. Ng, and A. Ortiz-Bobea. Temperature shocks and establishment sales. *The Review of Financial Studies*, 33(3):1331–1366, 2020.

E. Andreou and E. Ghysels. Sampling frequency and window length trade-offs in data-driven volatility estimation: appraising the accuracy of asymptotic approximations. In D. Terrell and T. B. Fomby, editors, *Econometric Analysis of Financial and Economic Time Series*, Advances in Econometrics, Volume 20 Part 1, pages 155–181. Emerald Group Publishing Limited, 2006.

F. M. Aragón, F. Oteiza, and J. P. Rud. Climate change and agriculture: Subsistence farmers' response to extreme heat. *American Economic Journal: Economic Policy*, 13(1): 1–35, 2021.

M. Arellano. *Panel Data Econometrics*. Oxford: Oxford University Press, 2003.

S. Arlot and A. Celisse. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79, 2010. doi: 10.1214/09-SS054.

M. Auffhammer, P. Baylis, and C. H. Hausman. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the united states. *Proceedings of the National Academy of Sciences*, 114(8):1886–1891, 2017.

A. Barreca, K. Clay, O. Deschenes, M. Greenstone, and J. S. Shapiro. Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the twentieth century. *Journal of Political Economy*, 124(1):105–159, 2016.

M. Burke and K. Emerick. Adaptation to climate change: Evidence from US agriculture. *American Economic Journal: Economic Policy*, 8(3):106–40, 2016.

M. Burke, S. M. Hsiang, and E. Miguel. Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239, 2015.

M. Burke, F. González, P. Baylis, S. Heft-Neal, C. Baysan, S. Basu, and S. Hsiang. Higher temperatures increase suicide rates in the united states and mexico. *Nature climate change*, 8(8):723–729, 2018.

C. Cattaneo and G. Peri. The migration response to increasing temperatures. *Journal of Development Economics*, 122:127–146, 2016.

G. Claeskens and N. L. Hjort. *Model Selection and Model Averaging*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2008. doi: 10.1017/CBO9780511790485.

X. Cui. Beyond yield response: Weather shocks and crop abandonment. *Journal of the Association of Environmental and Resource Economists*, 7(5):901–932, 2020.

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

T. Deryugina and S. Hsiang. The marginal product of climate. Technical report, National Bureau of Economic Research, 2017.

O. Deschênes and M. Greenstone. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American Economic Review*, 97(1):354–385, 2007.

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

S. Feng, A. B. Krueger, and M. Oppenheimer. Linkages among climate change, crop yields and Mexico–US cross-border migration. *Proceedings of the National Academy of Sciences*, 107(32):14257–14262, 2010.

M. Gammans, P. Mérel, and A. Ortiz-Bobea. Negative impacts of climate change on cereal yields: statistical evidence from France. *Environmental Research Letters*, 12(5):054007, 2017.

T. Garg, M. Jagnani, and V. Taraz. Temperature and human capital in india. *Journal of the Association of Environmental and Resource Economists*, 7(6):1113–1150, 2020.

E. Ghysels, P. Santa-Clara, and R. Valkanov. Predicting volatility: getting the most out of return data sampled at different frequencies. *J. Econometrics*, 131(1-2):59–95, 2006. ISSN 0304-4076. doi: 10.1016/j.jeconom.2005.01.004.

E. Ghysels, A. Sinko, and R. Valkanov. MIDAS regressions: further results and new directions. *Econometric Rev.*, 26(1):53–90, 2007. ISSN 0747-4938. doi: 10.1080/07474930600972467.

J. Graff Zivin, S. M. Hsiang, and M. Neidell. Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists*, 5 (1):77–105, 2018.

M. Harari and E. L. Ferrara. Conflict, climate, and cells: a disaggregated analysis. *Review of Economics and Statistics*, 100(4):594–608, 2018.

T. Hastie, R. Tibshirani, and R. Tibshirani. Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. *Statistical Science*, 35(4): 579 – 592, 2020. doi: 10.1214/19-STS733. URL `https://doi.org/10.1214/19-STS733`.

G. Heutel, N. H. Miller, and D. Molitor. Adaptation and the mortality effects of temperature across us climate regions. *Review of Economics and Statistics*, pages 1–33, 2017.

H. Hong and B. Preston. Bayesian averaging, prediction and nonnested model selection. *Journal of Econometrics*, 167(2):358 – 369, 2012. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2011.09.021. Fourth Symposium on Econometric Theory and Applications (SETA).

S. M. Hsiang. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proceedings of the National Academy of Sciences*, 107(35):15367–15372, 2010.

S. M. Hsiang, K. C. Meng, and M. A. Cane. Civil conflicts are associated with the global climate. *Nature*, 476(7361):438–441, 2011.

S. M. Hsiang, M. Burke, and E. Miguel. Quantifying the influence of climate on human conflict. *Science*, 341(6151):1235367, 2013.

K. Jessoe, D. T. Manning, and J. E. Taylor. Climate change and labour allocation in rural Mexico: Evidence from annual fluctuations in weather. *The Economic Journal*, 128(608): 230–261, 2018.

Y. Li, W. A. Pizer, and L. Wu. Climate change and residential electricity consumption in the yangtze river delta, china. *Proceedings of the National Academy of Sciences*, 116(2): 472–477, 2019.

R. Mendelsohn, W. D. Nordhaus, and D. Shaw. The impact of global warming on agriculture: a Ricardian analysis. *American Economic Review*, pages 753–771, 1994.

V. Mueller, C. Gray, and K. Kosec. Heat stress increases long-term human migration in rural pakistan. *Nature climate change*, 4(3):182–185, 2014.

R. G. Newell, B. C. Prest, and S. E. Sexton. The gdp-temperature relationship: implications for climate change damages. *Journal of Environmental Economics and Management*, page 102445, 2021.

R. J. Park, J. Goodman, M. Hurwitz, and J. Smith. Heat and learning. *American Economic Journal: Economic Policy*, 12(2):306–39, 2020.

J. T. Ritchie and D. S. Nesmith. Temperature and crop development. In J. Hanks and J. T. Ritchie, editors, *Modeling Plant and Soil Systems, Agronomy 31*, pages 5–29. American

Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 1991.

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

W. Schlenker, W. M. Hanemann, and A. C. Fisher. The impact of global warming on us agriculture: an econometric analysis of optimal growing conditions. *Review of Economics and statistics*, 88(1):113–125, 2006.

J. Shao. Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88(422):486–494, 1993. ISSN 0162-1459.

J. Shao. An asymptotic theory for linear model selection. *Statistica Sinica*, 7(2):221–264, 1997. ISSN 1017-0405. With comments and a rejoinder by the author.

C.-Y. Sin and H. White. Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, 71(1):207 – 225, 1996. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(94)01701-8.

E. Somanathan, R. Somanathan, A. Sudarshan, and M. Tewari. The impact of temperature on productivity and labor supply: Evidence from indian manufacturing. *Journal of Political Economy*, 129(6):1797–1827, 2021.

Q. H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2):307–333, 1989.

L. Wenz, A. Levermann, and M. Auffhammer. North–south polarization of european electricity consumption under future warming. *Proceedings of the National Academy of Sciences*, 114(38):E7910–E7918, 2017.

S. N. Wood. *Generalized additive models: an introduction with R.* CRC press, 2017.

Y. Yang. Can the strengths of aic and bic be shared? a conflict between model indentification and regression estimation. *Biometrika*, 92(4):937–950, 2005. ISSN 00063444. URL http://www.jstor.org/stable/20441246.

P. Zhang, J. Zhang, O. Deschenes, and K. Meng. Temperature effects on productivity and factor reallocation: Evidence from a half million Chinese manufacturing plants. *Journal of Environmental Economics and Management*, 88:1–17, 2018.

Table A1: Selecting Knots in Piecewise Yield Function Based on Minimized MSE

Unbalanced Panel				Balanced Panel			
One-knot Piecewise		Two-knots Piecewise		One-knot Piecewise		Two-knots Piecewise	
Knot	MSE	Knots	MSE	Knot	MSE	Knots	MSE
29	0.046905	24, 26	0.046478	30	0.029987	29, 33	0.047226
28	0.046959	23, 27	0.046483	31	0.030061	30, 32	0.047235
30	0.047258	25, 26	0.046488	29	0.030243	29, 34	0.047239
27	0.047331	24, 27	0.046497	32	0.030493	30, 31	0.047239
26	0.047903	31, 32	0.046499	28	0.030736	29, 32	0.047244
31	0.048043	30, 33	0.046499	33	0.031261	29, 35	0.047269
25	0.048573	22, 27	0.046505	27	0.031377	30, 33	0.047282
32	0.049233	23, 26	0.046507	26	0.032090	29, 31	0.047285
24	0.049271	30, 32	0.046536	34	0.032273	24, 26	0.047301
23	0.049950	25, 27	0.046538	25	0.032816	28, 36	0.047301

Notes: For illustration purpose, in each case, only the smallest ten MSEs and their corresponding knots are presented.

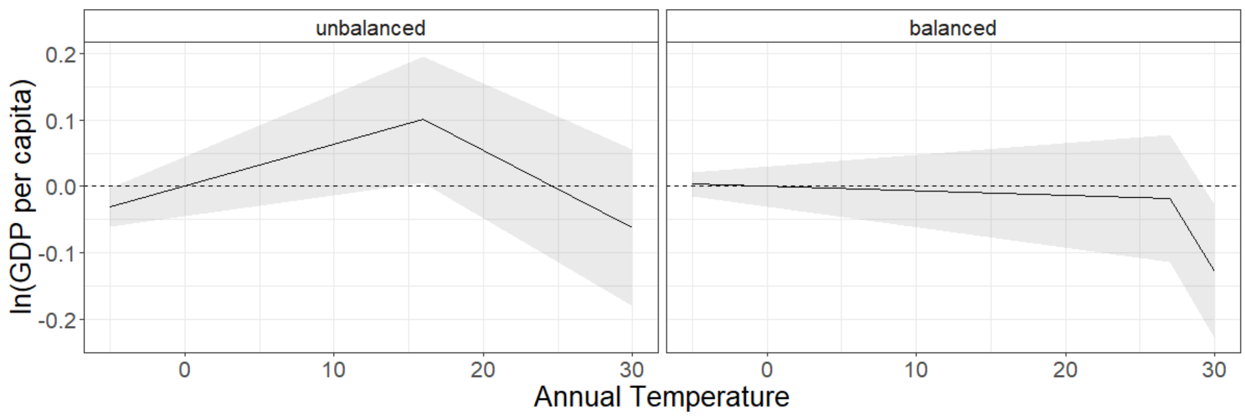


Figure A1: One-knot Spline Function of Temperature Effects on Economic Growth

Notes: Solid lines are spline functions recovered based on point estimates, and shallow bands are 95% confidence intervals constructed by applying the delta method on country-clustered standard errors.

Table A2: Weather Impacts on Corn Yields: Unbalanced Sample

	(1)	(2)	(3)	(4)
Average temperature: April	0.0110***			
	[0.0020]			
Average temperature: May	0.0033			
	[0.0029]			
Average temperature: June	-0.0093			
	[0.0057]			
Average temperature: July	-0.0586***			
	[0.0084]			
Average temperature: August	-0.0305***			
	[0.0036]			
Average temperature: September	0.0021			
	[0.0047]			
GDD (8-32C, in 100C)		0.0802*		
		[0.0316]		
GDD, squared		-0.0018*		
		[0.0007]		
HDD (34C+), squared root		-0.1240***		
		[0.0123]		
Degrees accumulated above 0C			0.0002***	0.0002
			[0.0001]	[0.0001]
Degrees accumulated above 24C				0.0059**
				[0.0017]
Degrees accumulated above 26C				-0.0107***
				[0.0022]
Degrees accumulated above 29C			-0.0056***	
			[0.0007]	
Precipitation	0.1794***	0.1141***	0.1055***	0.1068***
	[0.0169]	[0.0189]	[0.0201]	[0.0205]
Precipitation, squared	-0.0115***	-0.0087***	-0.0082***	-0.0084***
	[0.0012]	[0.0014]	[0.0015]	[0.0015]
Signal-to-noise ratio	0.2529	0.3813	0.4082	0.3935
Observations	120,995	120,995	120,995	120,995

Notes: Standard errors (in brackets) are state-clustered. Significance: * .05, ** .01, *** .001.

Table A3: Weather Impacts on Corn Yields: Balanced Sample

	(1)	(2)	(3)	(4)
Average temperature: April	0.0126***			
	[0.0021]			
Average temperature: May	0.0008			
	[0.0028]			
Average temperature: June	0.0014			
	[0.0038]			
Average temperature: July	-0.0453***			
	[0.0082]			
Average temperature: August	-0.0320***			
	[0.0028]			
Average temperature: September	0.0075			
	[0.0050]			
GDD (8-32C, in 100C)		0.1768***		
		[0.0284]		
GDD, squared		-0.0041***		
		[0.0007]		
HDD (34C+), squared root		-0.1491***		
		[0.0101]		
Degrees accumulated above 0C			0.0003***	0.0003***
			[0.0000]	[0.0000]
Degrees accumulated above 29C				-0.0055***
				[0.0007]
Degrees accumulated above 30C			-0.0099***	
			[0.0006]	
Degrees accumulated above 33C				-0.0084**
				[0.0024]
Precipitation	0.2245***	0.1498***	0.1380***	0.1347***
	[0.0293]	[0.0316]	[0.0294]	[0.0293]
Precipitation, squared	-0.0156***	-0.0123***	-0.0117***	-0.0114***
	[0.0025]	[0.0026]	[0.0025]	[0.0025]
Signal-to-noise ratio	0.3019	0.6194	0.6225	0.5688
Observations	44,814	44,814	44,814	44,814

Notes: Standard errors (in brackets) are state-clustered. Significance: * .05, ** .01, *** .001.

Table A4: Temperature Effects on GDP: Unbalanced Panel

						knot specification		
						0,5,...,25°C	0,3,...,27°C	16°C
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temp		-0.00042	0.01272***	0.01061***	0.01076***			
		[0.00206]	[0.00374]	[0.00298]	[0.00240]			
Temp, squared			-0.00049***	-0.00026	-0.00051			
			[0.00012]	[0.00033]	[0.00033]			
Temp, cubic				-0.00001	0.00002			
				[0.00001]	[0.00003]			
Temp, 4th order					0.00000			
					[0.00000]			
Temp splines:								
1st segment						0.01671***	0.01658***	0.00628*
						[0.00107]	[0.00109]	[0.00300]
2nd segment						0.00611	0.01055	-0.01161***
						[0.00419]	[0.00550]	[0.00320]
3rd segment						0.00370	0.00279	
						[0.00363]	[0.00313]	
4th segment						0.00946	0.00551	
						[0.00726]	[0.00441]	
5th segment						-0.01299**	0.00832	
						[0.00435]	[0.00818]	
6th segment						-0.00647	0.00579	
						[0.00497]	[0.00921]	
7th segment						-0.01389*	-0.01456*	
						[0.00571]	[0.00611]	
8th segment							-0.01283*	
							[0.00602]	
9th segment							0.00065	
							[0.00459]	
10th segment							-0.00589	
							[0.00588]	
11th segment							-0.03748***	
							[0.00681]	
Prec	0.01929	0.01894	0.01445	0.01449	0.01494	0.01568	0.01562	0.01412
	[0.01029]	[0.00996]	[0.00990]	[0.00992]	[0.01004]	[0.01015]	[0.01021]	[0.00997]
Prec, squared	-0.00566*	-0.00559*	-0.00475	-0.00476	-0.00486	-0.00500	-0.00498	-0.00466
	[0.00256]	[0.00253]	[0.00252]	[0.00252]	[0.00254]	[0.00256]	[0.00257]	[0.00251]
SNR	0.0009	0.0009	0.0046	0.0047	0.0048	0.0058	0.0084	0.0052
Observations	6,584	6,584	6,584	6,584	6,584	6,584	6,584	6,584

Notes: Standard errors (in brackets) are country-clustered. Significance: * .05, ** .01, *** .001.

Table A5: Selecting Knot in GDP Spline Based on Minimized MSE

Unbalanced Panel		Balanced Panel	
Knot	MSE	Knot	MSE
16	0.0026991	27	0.0022868
15	0.0026995	16	0.0022893
14	0.0027002	15	0.0022896
27	0.0027006	17	0.0022908
13	0.0027015	11	0.0022911
17	0.0027022	19	0.0022912
12	0.0027023	18	0.0022914
18	0.0027039	20	0.0022914
19	0.0027044	12	0.0022918
11	0.0027046	14	0.0022921

Notes: For illustration purpose, in each case, only the smallest ten MSEs and their corresponding knots are presented.

Table A6: Temperature Effects on GDP: Balanced Panel

						knot specification		
						5,. . . ,25°C	3,. . . ,27°C	27°C
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Temp		-0.0026	0.01019**	0.00574	0.01162			
		[0.00184]	[0.00329]	[0.00531]	[0.00774]			
Temp, squared			-0.00044**	-0.00004	-0.00098			
			[0.00014]	[0.00048]	[0.00116]			
Temp, cubic				-0.00001	0.00004			
				[0.00001]	[0.00006]			
Temp, 4th order					0.00000			
					[0.00000]			
Temp splines:								
1st segment						0.00488	0.01245*	-0.00067
						[0.00296]	[0.00526]	[0.00179]
2nd segment						0.00374*	0.0016	-0.03676***
						[0.00181]	[0.00171]	[0.00692]
3rd segment						0.00325	0.00541*	
						[0.00449]	[0.00263]	
4th segment						-0.00850*	0.00250	
						[0.00404]	[0.00265]	
5th segment						-0.00994	0.00517	
						[0.00575]	[0.00735]	
6th segment						-0.01306	-0.01030*	
						[0.00699]	[0.00470]	
7th segment							-0.01216	
							[0.00800]	
8th segment							-0.00212	
							[0.00330]	
9th segment							-0.00556	
							[0.00736]	
10th segment							-0.03803***	
							[0.00704]	
Prec	0.01606	0.01464	0.01019	0.0106	0.01079	0.01023	0.01057	0.01274
	[0.01134]	[0.01111]	[0.01093]	[0.01106]	[0.01111]	[0.01116]	[0.01134]	[0.01076]
Prec, squared	-0.00458	-0.00429	-0.00348	-0.00358	-0.00363	-0.00349	-0.00357	-0.0039
	[0.00297]	[0.00294]	[0.00291]	[0.00294]	[0.00295]	[0.00293]	[0.00297]	[0.00286]
SNR	0.0007	0.0011	0.0041	0.0043	0.0044	0.0043	0.0077	0.0054
Observations	4,300	4,300	4,300	4,300	4,300	4,300	4,300	4,300

Notes: Standard errors (in brackets) are country-clustered. Significance: * .05, ** .01, *** .001.