Journal of Environmental Economics and Management 108 (2021) 102445

Contents lists available at ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

The GDP-Temperature relationship: Implications for climate change damages

Richard G. Newell [a,b,*], Brian C. Prest [a], Steven E. Sexton [b]

[a] Resources for the Future, 1616 P St NW, Washington, DC, 20036, USA
[b] Duke University, 201 Science Drive, Durham, NC, 27708, USA

ARTICLE INFO

Article history:
Received 9 April 2020
Revised 23 February 2021
Accepted 14 March 2021
Available online 20 March 2021

Keywords:
Climate change
Econometrics
GDP impacts
Model uncertainty
Cross validation

ABSTRACT

Econometric models of temperature impacts on GDP are increasingly used to inform global warming damage assessments. But theory does not prescribe estimable forms of this relationship. By estimating 800 plausible specifications of the temperature-GDP relationship, we demonstrate that a wide variety of models are statistically indistinguishable in their out-of-sample performance, including models that exclude any temperature effect. This full set of models, however, implies a wide range of climate change impacts by 2100, yielding considerable model uncertainty. The uncertainty is greatest for models that specify effects of temperature on GDP *growth* that accumulate over time; the 95% confidence interval that accounts for both sampling and model uncertainty across the best-performing models ranges from 84% GDP losses to 359% gains. Models of GDP *levels* effects yield a much narrower distribution of GDP impacts centered around 1–3% losses, consistent with damage functions of major integrated assessment models. Further, models that incorporate lagged temperature effects are indicative of impacts on GDP levels rather than GDP growth. We identify statistically significant marginal effects of temperature on poor country GDP and agricultural production, but not rich country GDP, non-agricultural production, or GDP growth.

1. Introduction

1.1. Background and motivation

It has long been understood that economic outcomes are related to climate. This climate-economy relationship determines the scope and magnitude of market impacts from climate change over the next 100 years and beyond. Consequently, an understanding of the climate-economy relationship is central to projections of damages from anticipated climate change, and to policymaking that weighs the benefits and costs of climate change mitigation. Yet estimation of the scope and magnitude of climate impacts on the economy is hindered by the temporal invariance of climate over relevant time frames and by the correlation of cross-sectional climate variation with other regional heterogeneity that may effect economic performance, including historical effects of settlement and colonization (e.g., Acemoglu et al. 2002; Easterly and Levine 2003; Rodrik et al. 2004; Dell et al. 2014).

A recent literature, therefore, employs panel econometric methods to estimate the response of economic outcomes to *weather*, which is commonly defined as realizations from distributions of climatic variables, like temperature, wind, and precip-

* Corresponding author. Resources for the Future, 1616 P St NW, Washington, DC, 20036, USA.
E-mail addresses: newell@rff.org (R.G. Newell), prest@rff.org (B.C. Prest), steven.sexton@duke.edu (S.E. Sexton).

https://doi.org/10.1016/j.jeem.2021.102445
0095-0696/

itation (Dell et al. 2014; Hsiang 2016; Auffhammer 2018). This literature has estimated economically and statistically significant effects of weather on a variety of economic outcomes, including crop yields, industrial output, and labor productivity.[1] A subset of this literature also relates weather and weather shocks to economic aggregates like gross domestic product (GDP) (Hsiang 2010; Barrios et al. 2010; Anttila-Hughes and Hsiang 2011; Deryugina 2011; Dell et al. 2012; Hsiang and Narita 2012; Burke et al. 2015, 2018).

Much of this empirical research is intended to inform estimation of climate change damages and determinations of efficient climate change mitigation programs (Dell et al. 2014; Deryugina and Hsiang 2014; Hsiang 2016; National Academies of Sciences 2017; Burke et al. 2018). Integrated assessment models (IAMs), commonly used in analysis of climate change mitigation costs and benefits, rely upon the enumeration and aggregation of relevant, sector-specific impacts (National Academies of Sciences 2017). Given the sparseness of empirical estimates of sectoral impacts around the world, such models often must extrapolate impacts out of sample. Therefore, growing interest centers on the econometric estimation of climate impacts on economic aggregates that subsume sectoral effects, obviating the need to fully enumerate and estimate them. The aggregate econometric approach also complements the enumerative approach by potentially validating estimated magnitudes of damages. However, economic aggregates such as GDP are not direct welfare measures, and do not reflect non-market values affected by climate change. These should also be incorporated into welfare analysis.

Moreover, the econometric approach confronts two challenges in estimating aggregate economic impacts of climate change. First, identification of climate effects from weather variation requires strong assumptions about dynamic processes like adaptation and the persistence of idiosyncratic temperature responses amid secular climate change (Dell et al. 2014; Hsiang 2016). Second, theory does not prescribe specific, estimable, structural relationships between climate and economic outcomes (Dell et al. 2014; Hsiang 2016; Schlenker and Auffhammer 2018). Researchers, therefore, have made varying assumptions about the functional forms of these relationships.

For instance, by relating country-level, aggregate per capita economic growth to a log-linear function of temperature and precipitation, and controlling for country-specific effects and secular trends, Dell et al. (2012), henceforth DJO, estimated that only poor country growth is harmed by positive temperature shocks. In contrast to DJO, Burke et al. (2015), henceforth BHM, specified a quadratic relationship between temperature and per capita GDP growth that suggests rich and poor countries alike suffer from global warming and that both agricultural and industrial output growth are impeded. Employing parametric country-specific quadratic trends, the preferred model of BHM estimates a globally optimal temperature for GDP growth of 13 °C and predicts global income losses of 23% by 2100 due to unmitigated climate change. The same econometric approach is employed by Burke et al. (2018) to estimate a cumulative \$20 trillion in global damages avoided by 2100 if global warming is limited to 1.5 °Celsius (C) rather than 2 °C.

Whereas DJO and BHM each estimate a relationship between temperature and GDP *growth*, Hsiang (2010), Deryugina and Hsiang (2014) and Deryugina and Hsiang (2017) postulated a non-linear relationship between temperature and GDP *levels*. Hsiang (2010) relied principally upon a linear model to identify statistically and economically significant effects of annual average temperature on aggregate output and sectoral production in the Caribbean and Central America. His estimation of a piece-wise linear function relating daily average temperature to annual production indicated production losses occur only on extremely hot days with average temperatures of 27–29 °C. Specifying a similar piece-wise linear relationship between daily temperatures and GDP *levels*, Deryugina and Hsiang (2014) and Deryugina and Hsiang (2017) estimated U.S. production losses at daily average temperatures as low as 15 °C.[2]

Economists have long observed that theory often does not precisely define estimable forms of economic relationships, reserving to empiricists significant discretion in defining functional forms and selecting conditioning variables.[3] The consequences of that selection for inference have also long been enumerated.[4] As Hendry (1980) and Leamer (1983) observed, given that parameter sensitivity is indicative of specification error, empiricists often endeavor to demonstrate the robustness of parameter estimates to alternative assumptions. Yet, as Leamer (2010) contends, such sensitivity analyses or robustness checks are, themselves, often performed in *ad hoc* ways. Rigorous model selection tools can be applied in these settings to empirically ground model selection via processes less dependent upon researcher discretion.

1.2. Summary of methods and results

Thus, this paper systematically assesses the sensitivity of temperature parameter estimates to modeling assumptions and considers the implications of model uncertainty for estimates of climate change impacts on GDP. It evaluates competing models in the literature and a range of variants using a rigorous cross validation approach that is commonly employed in causal

[1] See for instance Deschênes and Greenstone (2007); Schlenker and Roberts (2009); Schlenker and Lobell (2010); Feng et al. (2010); Jones and Olken (2010); Lobell et al. (2011); Cachon et al. (2012); Fisher et al. (2012); Dell et al. (2014); Graff Zivin and Neidell (2014).

[2] Models relating economic outcomes to annual average temperatures are not directly comparable to those estimating responses to daily average temperatures given presumed non-linearities in the response function.

[3] For example, see Friedman (1953); Dhrymes et al. (1972); Cooley and LeRoy (1981); Leamer (1978, 1983); White (1996); Yatchew (1998); Hansen et al. (2011), and Belloni et al. (2014).

[4] See Keynes (1939), Koopmans (1947), Leamer (1978), Leamer (1983), Hendry et al. (1990), Chatfield (1996), and Sullivan et al. (1999), among others.

inference.[5] Models are evaluated according to out-of-sample model fit criteria as is particularly appropriate for models that are intended to predict future economic outcomes given expectations about future climatic conditions. Moreover, we invoke the substantial literature on data-driven, predictive model comparison (e.g., White 1996; Diebold and Mariano 1995; Hansen 2005; Hansen et al. 2011) to identify the set of models that are statistically superior to alternatives conditional on prediction procedures. This approach is standard (e.g., Diebold and Mariano 1995; West 1996), and was employed by Auffhammer and Steinhauser (2012) in the related context of modeling carbon emissions in the United States.

We use a country-level panel of economic growth, temperature, and rainfall to estimate the global relationship between GDP and temperature. Eight hundred models are estimated. They vary along four key dimensions: the assumed functional form for temperature and precipitation, methods of controlling for potentially confounding time trends, the persistence of temperature effects on GDP as indicated by the choice of GDP growth or levels as the relevant dependent variable, and the inclusion of temperature (and precipitation) lags as covariates. These models are evaluated by several cross-validation techniques to determine their relative performance, as well as their implications for damages from future warming.

Cross validation and statistical tests of model superiority reveal considerable model uncertainty that implies GDP impacts by 2100 ranging from substantial losses to substantial gains. Estimates of GDP impacts vary considerably more across those models assuming temperature effects on GDP *growth*, rather than GDP *levels*, reflecting the compounding of growth effects over time. For each cross validation approach, the set of superior models is dominated by levels models, but includes growth models. For superior *growth* models, the 95% confidence interval of GDP impacts in 2100 is −84% to +359%, reflecting considerable model and sampling uncertainty. In contrast, the 95% confidence region for superior *levels* models is −8.5% to +1.8%, and is centered around GDP losses of 1–2%. The model preferred by BHM that predicts GDP losses of 23% is excluded from all model sets of superior predictive ability.

Growth models yield immense uncertainty about global warming impacts. Across just those growth models that specify a non-linear temperature function, the combined model and sampling uncertainty yield a standard deviation of predicted impacts equal to 132% of GDP, with model uncertainty comparable in magnitude to sampling uncertainty.[6] Models specifying impacts on GDP levels, not growth, yield far less uncertainty in climate impacts; the standard deviation is equal to less than 3% of GDP for model, sampling, and combined uncertainty. Considerable growth model uncertainty affords little policy guidance and suggests caution is warranted when such estimates are incorporated into IAMs (e.g., Moore and Diaz 2015) or used to estimate the social cost of carbon (Ricke et al. 2018). Levels models, in contrast, are associated with less model uncertainty and project a range of impacts consistent with damage estimates embodied in leading IAMs.

Non-linear temperature specifications dominate the model sets of superior predictive ability in our prediction procedure. These include quadratic and cubic temperature functions, as well as temperature splines. These non-linear temperature models, however, do not perform statistically better in out-of-sample validation than models that exclude temperature entirely. In fact, the root mean-squared prediction errors of many of these models are not distinguishable to four decimal places, which reflects the relatively small share of variation explained by temperature. By forecast and backcast cross-validation approaches, models with any country-specific trends are statistically inferior to those that exclude trends, indicative of overfitting by models that include them.

Accounting for uncertainty among models of superior performance in our estimation procedure, we find that the marginal effect of temperature on GDP *growth* is not distinguishable from zero at annual average temperatures observed in our data. The marginal effect of temperature on GDP *levels* is more precisely estimated than the effect on GDP growth, yet there is still a wide range of temperatures for which the confidence intervals include zero.

We also explore temperature impacts on GDP by country-level income group and by agricultural versus industrial production. Among poor countries, we find consistently negative mean effects of temperature on GDP *levels*, which are statistically significant at the 10% significance level above 18 °C. Likewise, we find evidence of substantial temperature impacts on agricultural GDP levels, with a mean impact that is negative above 10 °C but statistically indistinguishable from zero effect at conventional levels of scientific certainty.

We find no statistically significant *growth* effects among poor countries or within the agriculture sector. Neither do we find statistically significant evidence of GDP level or growth effects among rich countries or non-agricultural production. These results are consistent with theories that industrialized countries with greater capacities to adapt to temperature and climate and economic sectors less exposed to weather and climate are less affected by climate change (Poterba 1993; Mendelsohn et al. 1994; Kahn 2005; Stern 2006; Nordhaus 2008; Tol 2009; Deryugina and Hsiang 2014).

The paper is organized as follows. The next section reviews the literature on the relationship between temperature and economic aggregates, highlighting the variety of modeling assumptions employed in the literature. Section 3 describes our method of assessing the impact of these alternative assumptions, and section 4 presents results of the model cross validation and implications for causal inference and climate damage projections. Section 5 concludes.

[5] For example, Friedman (1953); Varian (2014); Kleinberg et al. (2015); Christensen and Edward (2018), Athey (2017), and Athey et al. (2019).

[6] Growth models imply a range of impacts that is implausibly large given the overall historical exposure of the economy to temperature. The agricultural sector, for example, is one of the most exposed to climate change, but represents only a few percent of global income, whereas about two-thirds is services. Assumptions about the impact of discontinuous, catastrophic climate damages have been considered on the order of tens of percent (Kopits et al. 2014), but these types of impacts are not reflected in the historic data on which the GDP-temperature relationship is estimated.

2. Estimating economic responses to climate change

Research on agriculture, human capital, and other specific impacts of climate and temperature provide the microeconomic foundation for aggregate economic effects. These microeconomic foundations characterize a non-linear relationship between temperature and economic outcomes, with significant adverse production impacts occurring at daily average temperatures above about 29 °C.[7] It also demonstrates the alternative choices researchers have made in modeling the temperature relationship. The most flexible specifications of the micro-foundations literature use binned temperature observations, i.e., temperature step functions, to flexibly model non-linear relationships; but bin widths vary across these papers. Researchers have also documented impacts of temperature and climate on human health, conflict, and violence; factors that affect welfare, but less directly impact aggregate production.

The first estimates of the global welfare impacts of climate change were produced in the 1990s by combining assumptions about the extent of future warming with scientific evidence of its physical impacts and their valuation (Fankhauser 1994; Nordhaus 1994; Tol 1995, 2002a,b; Fankhauser 2013).[8] Recent statistical approaches to measuring aggregate economic impacts of climate relate observed economic outcomes to *weather* (i.e., short-term variations in climate), which varies in the cross-section and temporally, allowing the use of fixed effects to control for time-invariant region heterogeneity. This strand of the literature includes Hsiang (2010), DJO, Hsiang and Jina (2014), Deryugina and Hsiang (2014), BHM, and Burke et al. (2018). Though these papers all relate economic outcomes to temperature shocks, they differ in how they specify the equations used to estimate these responses. There are three principal model variations we explore: (1) GDP level effects, growth effects, and lag effects; (2) temperature functional form; and (3) controls for unobserved trends.

GDP Level Effects, Growth Effects, and Lag Effects. There is disagreement in the recent empirical literature as to whether temperature affects the level of economic output or its growth (e.g., see Schlenker and Auffhammer 2018). The modeling choice is not a trivial one. Growth effects compound over time, whereas level effects do not. Thus, and as we show, predictions of future losses from climate change vary considerably depending upon whether growth or level models are specified. The micro-foundations literature has largely related temperature to levels of economic outcomes, not their growth (Dell et al. 2012; Schlenker and Roberts 2009; Schlenker and Auffhammer 2018; Auffhammer 2018). The temperature impacts it documents, e.g., yield losses and reduced labor supply, are widely-accepted determinants of GDP. Some econometric models of economic aggregates relate temperature to GDP levels, as do the IAMs (National Academies of Sciences 2017; Hsiang 2010; Deryugina and Hsiang 2014). Yet other econometric models, including DJO, BHM, Hsiang and Jina (2014), and Burke et al. (2018) propose GDP growth may be affected beyond impacts on contemporaneous output.

The recent empirical literature does not provide theoretical foundations to favor models relating temperature to growth or levels. DJO characterize output as a multiplicative function of population, labor productivity, and exponentiated temperature. They offer the intuition that temperature may affect investment in institutions, which may affect productivity growth. BHM propose output is a function of temperature and total productive capacity, which depreciates over time and is rebuilt by savings. Savings, they assume, are permanently diminished during periods of high temperature and attendant low output. BHM also suggest that the rate of technological change is slowed by diminished cognitive capacity due to temperature change. These mechanisms are plausible, but have attracted little attention in the growth literature and have scant empirical support.

Likewise, there is little empirical record and little agreement on the interpretation of that record to advocate for GDP growth or levels specifications. DJO, for instance, report growth effects of lagged temperature. Whereas they interpret the coefficients on temperature lags as indicating growth effects, we contend sign reversal on temperature lags is indicative of a temporary

[7] See Schlenker and Roberts (2009); Ortiz-Bobea et al. (2018); Graff Zivin and Neidell (2014), and Sudarshan et al. (2015).

[8] This approach is embodied in three reduced-form IAMs that have informed the social cost of carbon in the United States (National Academies of Sciences 2017). This enumerative method contrasts with the statistical method employed by Mendelsohn et al. (2000a) and Mendelsohn et al. (2000b) and later Maddison (2003) and Nordhaus (2006) that inferred climate change costs from variation in economic activity across climates. Central estimates across these early studies range from a −11.5% loss of global GDP to a slight gain of 0.1% for temperature increases of 2.5–5.4 °C relative to pre-industrial levels (Tol 2014). Most models, including the IAMs used to estimate the U.S. social cost of carbon, predict losses of a few percent of GDP from a few degrees of warming (Greenstone et al. 2013; Tol 2014; Nordhaus and Moffat 2017). Whereas the enumerative approach of the IAMs accounts for non-market impacts and affords damage estimates that are traceable to specific mechanisms, it may overlook some channels by which climate change affects the economy, and it may extrapolate impacts out of sample to heterogeneous regions and production (Dell et al. 2012; Carleton and Hsiang 2016; National Academies of Sciences 2017).

effect on GDP levels.[9,10] Similarly, BHM estimate distributed lag models with 1–5 lags of a quadratic temperature polynomial to explore the persistence of temperature effects. In none of these BHM models is the cumulative temperature effect on growth statistically distinguishable from zero. Moreover, as in DJO, lagged temperature effects exhibit sign-reversal, implying transitory effects of temperature shocks on output.[11]

We incorporate into our following analysis models that include three lags of the contemporaneous temperature (and precipitation) function. As is later shown, some of these models exhibit out-of-sample performance that is statistically indistinguishable from some of the most accurate models excluding temperature lags. Lagged temperature effects on GDP growth are shown to be opposite sign and approximately equal in magnitude to contemporaneous effects further evidencing the transitory nature of temperature impacts and indicating level effects over growth effects.

Temperature Functional Form. The second dimension along which the climate-economy literature varies is specification of the function relating temperature to economic outcomes. This choice also dramatically affects the magnitude of damage estimates. The preferred model of DJO specifies a linear temperature effect, implying that a temperature shock affects economic outcomes similarly regardless of the mean from which temperatures deviate.[12] Like Hsiang (2010) and Deryugina (2011), DJO also implement more flexible, piece-wise linear functions of temperature that accommodate asymmetric effects of small increases and decreases in temperature relative to an optimum.[13]

BHM favor a quadratic relationship between temperature and growth that allows warming to boost growth in countries with cold climates and impede growth in hot countries. Using data substantially similar to DJO, they estimate statistically and economically significant growth effects of temperature shocks in rich and poor countries alike, and across both agricultural and industrial production.[14] The quadratic specification estimated by BHM identifies an optimal annual average temperature for GDP growth of 13 °C from which deviations in either direction generates changes in growth of equal magnitude but opposite sign. The quadratic temperature relationship is more flexible than the linear relationship specified by DJO. Yet it imposes a symmetry of growth effects due to temperature deviations away from the optimum that abstracts from the micro-foundations evidence.[15] BHM also consider higher-ordered polynomials of temperature, as well as restricted cubic splines with 3–7 knots.[16]

Controls for Unobserved Trends. The third major dimension of model heterogeneity within the existing literature is the choice of controls for trending unobservables. Theory offers little guidance in controlling for trending unobservables, and the extant literature appears to take a fairly *ad hoc* approach to modeling trend heterogeneity. Because of heterogeneity in endowments, institutions, and history, countries or regions are likely to have varying growth capabilities. The parametric trends employed

[9] This is particularly true given serial correlation in country average annual temperatures that implies a relatively hot year is typically followed by another relatively hot year, e.g., due to El Nino-Southern Oscillation and other decadal oscillations (Hsiang 2010). In BHM's data, the first lag and contemporaneous temperature have a correlation coefficient of 0.5. This correlation is depicted for each country in Appendix Fig. A1. Contemporaneous temperature and more distant lags are also positively correlated, though, as expected, the correlation declines with lag distance.

[10] DJO interpret their lagged-effects models differently than we do because they assume a transitory contemporaneous temperature effect on GDP growth *and* a temperature effect on GDP level. See their equations (2)-(3). This assumption is strong, and seems counter to the intuition they provide—that temperature may affect institutions that affect growth. If investment in institutions is low relative to some counterfactual during a temperature shock, then institutional investment remains low indefinitely in subsequent periods unless an offsetting temperature shock occurs. Thus, a temperature shock that affects GDP growth should affect growth in subsequent periods, producing lagged temperature effects that exhibit common sign and magnitudes to the contemporaneous effect. For example, if we drop their levels effect entirely (set $\beta = 0$) but allow for a lagged effect on productivity growth ($\gamma_2 < 0$) in addition to the existing contemporaneous effect ($\gamma_1 < 0$), the growth rate equation would become $g_{it} = g_i + \gamma_1 T_{it} + \gamma_2 T_{it-1}$; this result features a common sign on contemporaneous and lagged temperature, contrary to DJO's results in which the signs differ. The only conceptual ways to find differing signs as DJO estimate is with the existence of a levels effect, or, more perversely, a growth effect that reverts itself with one lag. Further, if there are no levels effects, then the sum of all temperature coefficients should be at least as large as the magnitude of the contemporaneous effect. Yet none of the lagged temperature coefficients in DJO models is statistically significant at conventional levels. All are small in magnitude, and the sum of coefficients is smaller in magnitude than the contemporaneous effect. All of this suggests temperature effects on levels rather than growth, contrary to the interpretation of DJO.

[11] BHM do not report these coefficients in their main text or supplementary information. We produced the coefficients and their standard errors using BHM replication data and code. The addition of temperature lags to the models that perform best in our model cross validation also yields little evidence of persistent temperature effects.

[12] The model implies that contemporaneous growth among poor countries declines by 1.4 percentage points annually for each 1 °C of warming.

[13] Cautioning against over-interpretation due to data reliability concerns, they report coefficients that characterize approximately linear effects that are statistically indistinguishable from zero for poor countries across all temperature bins. For rich countries, temperature coefficients are approximately zero and not statistically significant, except in the range of 15–25 °C, within which coefficients are positive and barely significant.

[14] BHM impose a globally quadratic relationship, not to be confused with a within-country quadratic relationship. As shown by McIntosh and Schlenker (2006), these two assumptions are conceptually different on a fundamental level, and, therefore, have significant practical implications. For example, a global quadratic implies a single centering point (here, "GDP-maximizing temperature") for all countries and years, compared to many country- and year-specific centering points.

[15] BHM state conditions under which the annual aggregation of daily temperatures used in some of the micro-foundations literature (e.g., Schlenker and Roberts 2006; Graff Zivin and Neidell 2014; Sudarshan et al. 2015; Stevens 2017) yields a temperature response curve that is concave, smoother, and characterized by a lower optimum temperature than the micro responses. It is also important to note that some of the micro-foundations literature relates outcomes to daily maximum temperature, which is characterized by a higher mean than daily, monthly, or annual average temperatures that incorporate temperature readings from relatively cool nighttime periods. However, the concave relationship BHM define in their equation (1) does not impose symmetry. See their Fig. 1(f).

[16] These alternative specifications all confirm a concave relationship between temperature and growth, but the peaks of the concave relationships vary across specifications in non-trivial ways, typically implying GDP-maximizing temperatures greater than 13 °C.

Table 1
Model assumptions of climate-economy literature.

	Model Specifications					
	Dependent Variable	Temperature Function	Country FEs	Year FEs	Region X Year FEs	Time Trend Polynomial
Dell et al. (2012)	Growth	Linear	•	•	•	
Burke et al. (2015)	Growth	Quadratic	•	•		3
Hsiang (2010)	Levels	Piecewise linear	•	•	•	2
Deryugina and Hsiang (2014)	Levels	Piecewise linear	•	•		
Hsiang and Jina (2014)	Growth	NA	•	•		2
Burke et al. (2018)	Growth	Quadratic	•	•		3

Notes: Table includes key modeling assumptions for select studies of the climate-economy literature. Dependent variables are either GDP levels or GDP growth. Time trend polynomials are of GDP levels.

by BHM and Hsiang and Jina (2014) permit country-level heterogeneity, but constrain the functional form of country trends.[17] Such parametric trends can result in over-fitting, as we demonstrate they do in this setting.[18]

Fixed effects, in contrast, flexibly and non-parametrically control for trends, but they do not admit country-specific trends.[19] Models like DJO that are saturated with fixed effects lend credible causal inference as they are robust to many sources of omitted variables bias, but they may also absorb variation necessary to identify some relationships (e.g., Fisher et al. 2012; Deschênes and Greenstone 2012).[20],[21] In the present context, saturation of fixed effects or parametric time trends can both lead to this problem.[22]

Table 1 summarizes the varying assumptions of prominent papers in the climate-economy literature along key dimensions.[23] The cross validation exercise described in the subsequent section considers the sensitivity of temperature parameters to modeling assumptions along these dimensions.

3. Data and methods

Given the model uncertainty evident in the growing climate econometrics literature, empiricists must make choices about functional forms and inclusion or exclusion of controls. Such model ambiguity is important to the extent that outcomes of interest differ markedly across alternative empirical models, yielding substantial model uncertainty. In the following sections, we assess the performance of alternative models and the magnitude of model uncertainty by employing cross validation in the spirit of Athey (2017). Similar to Athey et al. (2019), which introduces ensemble methods to causal inference, alternative models are evaluated according to their out-of-sample predictive abilities. This approach is also conceptually similar to that of Auffhammer and Steinhauser (2012), which considers model uncertainty in CO_2 emissions forecasts.

3.1. Model specifications

The degree to which the data recommend a particular relationship between temperature and GDP is evaluated using standard model cross-validation techniques that fit the parameters of specific models using only a subset of available historical data. The predictive performance of alternative models is then assessed on the remainder of the historical data. We consider the performance of the models preferred by BHM and DJO, the only other global econometric assessments of the GDP-temperature relationship, as well as 799 variants that incorporate alternative functional forms for the temperature and precipitation responses and varying controls for unobserved trends. We also evaluate models that reflect alternative assumptions about the persis-

[17] BHM use country fixed effects to control for time-invariant country heterogeneity and also use a set of year fixed effects to control for global trends in growth. Rather than controlling for regional or country-type (e.g., poor) trends non-parametrically as in DJO, BHM introduce a parametric country-specific quadratic time trend. Because the dependent variable in their regressions, growth, is the first derivative of income, their quadratic trend implies a country-specific cubic polynomial in income levels. BHM report that estimation results look similar with only a linear trend in growth. However, as we show in this paper, estimated GDP impacts vary considerably across alternative specifications. Hsiang and Jina (2014) include a linear country-specific time trend in estimating the relationship between economic growth and cyclone exposure in the Caribbean. This imparts a quadratic time effect in levels of production, similar to Hsiang (2010). They also include year fixed effects to flexibly control for common trends and country fixed effects to control for time-invariant heterogeneity.

[18] Removing the country-specific quadratic time trends from BHM's specification as well as adding them to DJO's specification changes the sign of the estimated impacts of warming on GDP in 2100, as shown in Table A1.

[19] Deryugina and Hsiang (2014) exclude parametric time trends in estimating the production responses of U.S. counties as a flexible function of temperature. They use county and year fixed effects to control for common trends and time-invariant county heterogeneity.

[20] DJO employ country fixed effects, year fixed effects interacted with regional indicators, and year fixed effects interacted with an indicator for countries that are poor when they enter the data. "Poor" is defined as per capita GDP below the country median at the earliest period recorded. This saturates the model with fixed effects to non-parametrically control for unobservables. It is robust to region-specific time trends that might be unique to rich or poor countries.

[21] As discussed by Fisher et al. (2012), oversaturation of fixed effects can amplify attenuation bias in the presence of measurement error. Too many controls will absorb most of the variation in the data, leaving measurement error to play a larger role in the remaining identifying variation.

[22] DJO include 300 region-year fixed effects that BHM do not, whereas BHM include 332 time trend variables not included in DJO.

[23] Hsiang (2010) includes country and industry-specific quadratic time trends in his models of the production responses to temperature change in 28 Caribbean-basin countries, as well as industry-region-year and industry-country fixed effects. Because GDP growth is the first derivative of income levels, the quadratic time trend in Hsiang (2010) is analogous to a linear time trend in a growth model.

tence of temperature effects on GDP, that is, whether temperature affects GDP levels or growth, and whether temperature and precipitation lags affect outcomes.

The specifications we consider take the following form for country i in region r during year t:

$$\Delta \ln(GDP_{i,r,t}) = \beta_0' X_{i,t} + \beta_1' X_{i,t-1} + \beta_2' X_{i,t-2} + \beta_3' X_{i,t-3} + \beta_4' X_{i,t-4} + \alpha_i + \lambda_{r,t} + h_i(t) + \varepsilon_{i,t}, \quad (1)$$

where $GDP_{i,r,t}$ is GDP per capita, α_i are country fixed effects, $\lambda_{r,t}$ are alternative time fixed effects, and $h_i(t)$ are alternative time trends. The $X_{i,t}$ term is a vector representing functions of the levels of temperature and precipitation. Focusing on the temperature component of $X_{i,t}$, this includes functions of annual temperature that are alternatively excluded or specified as a linear, quadratic, or cubic polynomial, or as a cubic spline in temperature (and analogously for precipitation). The $X_{i,t-\ell}$ terms are ℓ-order lags of these functions.[24] The temperature variable in $X_{i,t}$ measures average annual temperature in degrees C. The precipitation function is similarly varied across specifications, where the primitive variable measures cumulative annual rainfall in millimeters. An idiosyncratic error, $\varepsilon_{i,t}$ is permitted to be correlated within countries across time and within years across countries.[25]

Time fixed effects (λ), which control for trending unobservables, are modeled in two ways. First, we consider simple year fixed effects (as in BHM), which allow for secular trends, but do not account for country or region-specific trends in GDP growth. Second, we consider more flexible region-year fixed effects (as in DJO and Hsiang, 2010), which admit distinct GDP growth shocks across regions.[26] Identification comes from intra-region variation in temperature shocks. Country-specific time trends are also included in some specifications, including polynomials of degree 1–3 for $h_i(t)$. BHM prefer quadratic trends in growth, which constitute cubic trends in GDP, while Hsiang (2010) and Hsiang and Jina (2014) adopt linear growth trends. All models that we evaluate include the α_i country fixed effects.[27]

Because GDP (in levels) is a non-stationary series, proper inferences over a levels effect requires that we specify a GDP-level response to temperature in first differences. This specification can be derived as a special case of equation (1).[28] Hence, the levels-versus-growth distinction can be interpreted as special cases of the β_ℓ parameters in equation (1). For example, if the response of (log) GDP levels is thought to be a quadratic function of contemporaneous temperature (and not its lags), then denoting temperature as $x_{i,t}$ (so $X_{i,t} = (x_{i,t}, x_{i,t}^2)$) we have

$$\begin{aligned}\ln(GDP_{i,r,t}) &= \beta_0' X_{i,t} + \cdots \\ &= \beta_{0,1} x_{i,t} + \beta_{0,2} x_{i,t}^2 + \cdots,\end{aligned}$$

then the first-differenced, stationary, and estimable relationship is:

$$\begin{aligned}\Delta \ln(GDP_{i,r,t}) &= \beta_{0,1} \Delta x_{i,t} + \beta_{0,2} \Delta[x_{i,t}^2] + \cdots \\ &= \beta_0' \Delta X_{i,t} + \cdots \\ &= \beta_0' X_{i,t} - \beta_0' X_{i,t-1} + \cdots,\end{aligned}$$

which uses growth rates (log-differences of per capita GDP) as the dependent variable. The quadratic differenced temperature term is the change in temperature-squared, which we note is conceptually very different from the squared change in temperature, i.e., $(\Delta x_{i,t})^2$. As the last two equations show, this data-generating process can be estimated by regressing growth on the

[24] We use four-knot splines, with knots placed at the four interior ventiles of the temperature data: (11 °C, 18 °C, 23 °C, and 26 °C), corresponding to the 20th, 40th, 60th, and 80th percentiles of annual temperature observations. The knots for the spline in precipitation are defined analogously.

[25] Following BHM and DJO, and except where specified, errors are clustered by country and year or region-year corresponding to the fixed effects specified in the models. A substantial literature investigates the time series properties of global surface temperatures, and though consensus is thus far elusive, we note that Kaufmann et al. (2010) concludes that surface temperature shares a stochastic trend with radiative forcing, and that this implies correlation of errors across years. We investigate the implication of this correlation for the standard errors on temperature coefficients in our regression models. We find that standard errors tend to increase when clustering across time in five-year blocks. P-values are not substantially different, and inferences are virtually unchanged. These results are available from the authors upon request.

[26] We adopt the DJO specification of regions, as follows: Middle East/North Africa, Sub-Saharan Africa, Latin America and Caribbean, Western Europe and offshoots, Eastern Europe and Central Asia, and Asia and Pacific Islands. See the DJO appendix for a full list of countries.

[27] Estimation with fixed effects controls flexibly for unobservables that may confound identification of temperature effects. The inclusion of fixed effects permits identification only from deviations from group-specific means. That is, we identify effects of temperature changes *relative* to, for instance, country-specific means. The inclusion of fixed effects precludes identification of responses to *absolute* temperatures. Though we do not include in our main analysis any models that omit country or year fixed effects, we nevertheless explore the sensitivity of marginal temperature effects to their inclusion of exclusion. For each model in the union of model confidence sets, we estimate the difference in marginal temperature effects between baseline models that include fixed effects and the analogous models that exclude fixed effects. These differences are estimated for 1000 bootstrapped samples for each model, yielding confidence regions depicted in Fig. A4. As shown, we do not identify statistical differences across models that include or exclude the FEs of the baseline models at even the 50% significance level. Because of the potential bias in estimated temperature effects from models that exclude fixed effects, we are hesitant to conclude from this analysis that absolute and relative temperatures affect economic outcomes similarly.

[28] While BHM (and DJO) note the non-stationarity concern and, hence, the need to take first differences, they only take first differences in the left-hand side of the equation, GDP, without taking first differences in the right-hand side (temperature and precipitation). Consequently, their estimated models are not directly derived from their conceptual models.

corresponding first-difference of the specified temperature response, i.e., $\Delta X_{i,t}$, or equivalently, by using specification equation (1) under the constraint that $\beta_1 = -\beta_0$ and $\beta_2 = \beta_3 = \beta_4 = 0$.

The models preferred by BHM and DJO are also variants of equation (1). BHM regress log-GDP growth on a quadratic of temperature and precipitation, year and country fixed effects, and country-specific quadratic trends. DJO specify log-GDP growth as a function of temperature and country, region-year, and poor-year fixed effects.[29]

In total, we evaluate 800 possible specifications resulting from the following modeling choices and parameter restrictions in equation (1):

1. Terms in the weather vector $X_{i,t}$:
   - Temperature function (5 forms): excluded, 1–3° polynomials, or spline;
   - Precipitation function (5 forms): excluded, 1–3° polynomials, or spline;
2. GDP growth versus level effect (4 forms):
   - Growth, no lags: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$;
   - Growth, with three lags: $\beta_4 = 0$;
   - Level, no lags: $\beta_0 = -\beta_1$ and $\beta_2 = \beta_3 = \beta_4 = 0$;
   - Level, with three lags: $\beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$;[30]
3. Time and region controls (8 forms):
   - Time fixed effects (2 forms): simple year (BHM-style) or region-year (DJO-style); and
   - Country-specific time trends (4 forms): none and 1–3° polynomials.[31]

3.2. Data

Models are estimated using the same data employed by BHM. Specifically, we use the 2012 World Development Indicators (World Bank 2012) country-year panel of real annual GDP per capita for 166 countries from 1960 to 2010. The data include 6584 country-year observations.[32] Also as in BHM, we use Matsuura and Willmott (2012) gridded, population-weighted average temperature and precipitation data aggregated to the country-year level. Because we estimate a subset of models with three temperature (and precipitation) lags and because estimation is constrained by the historical extent of the GDP series, estimation proceeds over approximately 4% less data than in BHM's main specification.

While only GDP, temperature, and precipitation data are used for estimation, we also forecast the GDP impacts of the alternative parameter estimates using the same method as BHM. As in BHM, projections of population and economic growth are drawn from the Shared Socioeconomic Pathways (O'Neill et al. 2014) scenario 5 (SSP5). For comparison to BHM, we use the representative carbon pathway RCP8.5 as a benchmark scenario of unmitigated future warming (van Vuuren et al. 2011). It represents the ensemble average of all global climate models contributing to CMIP5, the Coupled Model Intercomparison Project phase 2010–2014 that informed the fifth assessment report of the Intergovernmental Panel on Climate Change.[33] RCP8.5 corresponds to an expected increase of 4.3 °C in global mean surface temperature by 2100 relative to pre-industrial levels (Stocker et al. 2013).

3.3. Cross validation techniques

Given theoretical ambiguity about which econometric models correctly capture the data generating process, model performance is a useful criterion for model selection. As Hsiang (2016) notes, the parameters empirically recovered in climate econometric models are "put to work" in order to "inform projections of future outcomes under different climate scenarios."[34] Consequently, the out-of-sample prediction properties of models are arguably the performance characteristics of primary importance.

In-sample fit criteria are prone to selecting over-fitted models, particularly as high-ordered polynomials of covariates are introduced in some models (Chatfield 1996). Commonly reported statistics, like adjusted R^2, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) assess in-sample model fit with parametric penalties for the inclusion of weakly cor-

---

[29] Poor-year fixed effects are interactions of year indicators and an indicator for whether a country was poor when it entered the data, defined as per capita GDP below the median.

[30] A model where the level of GDP is affected by $X_{i,t}$ corresponds to a first-differenced model with three lagged first differences, which in turn corresponds to a case of Eq. (1) where the sum of the levels effects nets to zero: $\alpha_0' \Delta X_{i,t} + \alpha_1' \Delta X_{i,t-1} + \alpha_2' \Delta X_{i,t-2} + \alpha_3' \Delta X_{i,t-3} = \alpha_0' X_{i,t} + (\alpha_1 - \alpha_0)' X_{i,t-1} + (\alpha_2 - \alpha_1)' X_{i,t-2} + (\alpha_3 - \alpha_2)' X_{i,t-3} - \alpha_3' X_{i,t-4}$.

[31] All time trends are referenced according to the order of the polynomial as it would appear in the growth model, i.e., Eq. (1). Hence, a "linear" time trend enters a GDP growth equation as a linear trend and corresponds to a quadratic time trend in GDP levels.

[32] The panel is not balanced because the data series is not complete for the full period for some countries. In particular, some countries are added to the series post-1960.

[33] See http://cmip-pcmdi.llnl.gov/.

[34] For instance, such parameter estimates can be used to generate estimates of the net present value of future damages from emitting a ton of greenhouse gases. These estimates of the social cost of carbon are used to evaluate the benefits of carbon reductions (National Academies of Sciences 2017).

related covariates. Each has flaws (Green 2012). Therefore, we employ model cross-validation (CV) to assess the performance of alternative models of the GDP and temperature relationship. By training the model over a subset of the data (the training, or estimation set), and assessing its predictive accuracy on a hold-out sample (the test set), the CV approach is expressly non-parametric and avoids the *ad hoc* penalties of alternative, in-sample measures of model performance (Stone 1974; Snee 1977).[35] Out-of-sample validation is not a new approach, and the literature has a rich history (e.g., Diebold and Mariano 1995; West 1996). In a study related to climate change specifically, Auffhammer and Steinhauser (2012) used such an approach to evaluate competing models for CO_2 emissions forecasts. Athey et al. (2019), likewise, use out-of-sample predictions to assign model weights in an ensemble approach to causal inference in panel data settings characterized by model ambiguity.

We undertake cross validation of 800 models using three CV methods that divide data into distinct training and test sets. These three methods are: forecast, backcast, and K-fold CV. Forecast CV is implemented by dividing the historical data into training sets of early data and testing on later data.[36] This is a standard and intuitive approach for evaluating the out-of-sample performance of statistical models of time series data (e.g., see Raftery et al. 2017 and Athey et al. 2019). Backcast CV is similar to forecast CV, but implemented by training on the most recent data and testing on early data.[37] Both forecast and backcast CV approaches explicitly account for the time-series nature of the data. In contrast, K-fold CV proceeds by dividing the data randomly into K groups and iteratively training the model K times on $\frac{K-1}{K}$ of the data. The model estimated in each iteration is tested on the remaining $\frac{1}{K}$ of the data.[38] We implement K-fold CV for completeness and to avoid exercising researcher discretion, even though K-fold CV ignores the time-series nature of the data and yields an optimistic estimate of model fit if data are serially correlated. K-fold CV also does not extrapolate out of sample where over-fitted models perform poorly.

When econometric models include time fixed effects, cross validation methods do not generate parameter estimates for fixed effects appearing in test data that do not appear in training data. For example, a model estimated on 1981–2000 data will have no estimated fixed effect for the year 2001 that appears in a forecast CV test set.[39] Auffhammer and Steinhauser (2012) address this problem by estimating a linear trend in the fixed effects of the trained model and predicting on test data using extrapolated fixed effects.

We adopt a different approach. For forecast and backcast CV, we remove fixed effects from the models prior to estimation by demeaning both the dependent and explanatory variables. This results in precisely the same coefficient estimates on the remaining parameters, such as the coefficients on temperature. We demean GDP growth and all explanatory variables (temperature, temperature squared, precipitation, precipitation squared, the country-specific time trend polynomials, etc.) by the fixed effect groups. This involves demeaning by both country and year or region-year before implementing the CV.[40] Because we remove more variation from the data prior to estimation for those models that include region-year fixed effects, we implement the CV procedures separately for these models. Consequently, we do not assess whether relative performance differences of models with year fixed effects versus region-year fixed effects are statistically different under the forecast or backcast CV approaches. However, we can do so in K-fold CV where estimating year fixed effects is feasible.

In summary, each model is estimated iteratively on separate training sets, generating estimated model parameters governing the temperature, precipitation, and time trend functions in Eq. (1). Then predicted values for the test set are computed by applying these estimated parameters to the observed independent variables in the test set (data which were not used to estimate the model) as follows

$$\text{Predicted Demeaned Growth}_{i,r,t} = \hat{\chi} + \sum_{\ell=0}^{4} \hat{\beta}_{\ell} X_{i,t-\ell} + \hat{h}_i(t) \tag{2}$$

[35] See also Arlot and Celisse (2010) for a more recent survey.

[36] We implement forecast CV rolling estimation windows of approximately 20 years: 1964–1985, 1971–1990, 1976–1995, 1981–2000, and 1986–2005. Test windows of five years immediately follow each estimation window. (The first window, 1964–1985, is slightly longer than 20 years to hold the test window fixed at five years across all implementations of the CV.) The fixed, rolling estimation windows are employed in order to implement the MCS procedure for nested models. See Hansen et al. (2011) and Elliott and Timmermann (2016). Short test windows are deliberately employed to be conservative in discerning the performance of flexible-time-trend models. Even amid short test windows, these models perform poorly relative to models that exclude flexible trends. Longer test windows would exacerbate their prediction errors. Athey et al. (2019) also use short forecast windows. For all cross-validation methods, we drop countries with fewer than 10 years of data in the training set to avoid bias in the specification of flexible time trends.

[37] Backcast CV estimation windows are 1969–1988, 1974–1993, 1979–1998, 1984–2003, and 1989–2010. Test windows of five years immediately precede each the first year of each estimation window. (The last window, 1989–2010, is slightly longer than 20 years to hold the test window fixed at five years across all implementations of the CV.)

[38] We set $K = 5$, as is common in the literature. Thus, a given model is estimated each time using 4/5 of the data, and its accuracy is tested on the remaining 1/5 (Geisser 1975). This process is repeated 5 times, so that every observation is used in a test set exactly once. K-fold CV is conceptually similar to leave-one-out CV, but less computationally intensive. Using multiple splits avoids the randomness inherent in splitting the data only once (Opsomer et al. 2001). The random sampling to divide the observations is not block sampled because doing so would make it impossible to estimate certain coefficients such as country-specific time trends. In the event a training set incorporates no observations for a given country, observations for that country are omitted from the test set.

[39] This is primarily a problem for forecast and backcast approaches; in K-fold CV individual years are not systematically dropped from the training set, and all parameters can be estimated.

[40] Because the panel is not balanced, simply demeaning first by country and then by year does not produce a demeaned dataset. Therefore, we demean the data by employing the method of alternating projections. In particular, we use the `demeanlist()` function from the R package `lfe`. Importantly, all polynomials and spline bases are computed prior to the demeaning process, including the first difference of these bases for the GDP levels specification. Failure to do so would result in different coefficient estimates for the temperature relationship (among others) under the demeaning and indicator variable approaches.

The estimated constant $\widehat{\chi}$, weather coefficients $\widehat{\beta}_{\ell}$, and country-specific time trend functions $\widehat{h}_i(\cdot)$ all vary across model specifications and training sets. The left-hand side is labelled predicted demeaned growth because that is the dependent variable in the model fit (that is, after the removal of the fixed effects: Demeaned $\text{Growth}_{i,r,t} = \Delta \ln(GDP_{i,r,t}) - \widehat{\alpha}_i - \widehat{\lambda}_{r,t}$).[41] The out-of-sample accuracy of this prediction is then evaluated by calculating its root mean square error relative to observed demeaned growth.

3.4. Model confidence sets

Given the common CV performance across many models that we demonstrate in Section 4, a determination of which models are statistically significantly superior to alternatives in their predictive ability is not obvious. Therefore, and given *ex ante* theoretical ambiguity about estimable relationships, we employ the model confidence set (MCS) procedure of Hansen et al. (2011), which iteratively eliminates from consideration models that are inferior to alternatives in their predictive ability at the 95% confidence level. Models remaining under consideration are statistically indistinguishable in their performance. The procedure considers a null hypothesis that model losses (prediction errors) are equivalent across models. If the null is rejected, an elimination rule removes a model from consideration and the null is tested again. The procedure iterates until the equivalence of model losses cannot be rejected. Because this procedure explicitly compares all models simultaneously, it is more extensive and comprehensive than traditional approaches to model selection, such as running an F-test for each model individually (essentially comparing each individual specification to a null specification) or running a Davidson-MacKinnon J-test (which considers pairwise model comparisons).

Like Hansen et al. (2011), we employ a decision rule that iteratively eliminates the model with the greatest standardized loss relative to the *average* of models remaining in the consideration set, denoted by the test statistic $T_{\max,\mathcal{M}}$. Because we likely remove more variation from those models with more saturated fixed effects due to the demeaning procedure, i.e., the region-year models, such models are likely advantaged in tests of predictive ability. Hence, forecast and backcast CV MCSs are separately estimated for models that include year or region-year fixed effects. We cannot determine whether differences in the relative performance of models along this dimension are statistically significant in forecast and backcast CV.

3.5. Projections of GDP impacts of climate change

We project the impact of expected warming on global GDP by 2100 using the RCP8.5 climate projection as a benchmark of unmitigated climate change and SSP5 for projections of moderate to strong baseline GDP and population growth, similar to BHM (O'Neill et al. 2014). Neither we nor BHM forecast baseline GDP using the econometric model. Country-level projections of economic growth, population growth, and climate warming are combined with the estimated relationship between GDP growth and temperature to predict changes in future growth rates for each country and year. Given the estimated concave growth-temperature relationship, baseline GDP growth is incremented in cold countries as warming occurs and decremented in hot countries. We employ the methodology and data of BHM's preferred projection. For a more detailed description of this projection, see section D of BHM's supplementary information.

Following BHM, we allow per capita GDP to evolve according to:

$$GDP_{i,t} = GDP_{i,t-1} \times \left(1 + \eta_{i,t} + \delta_{i,t}\right),$$

where $\eta_{i,t}$ is the economic growth rate absent temperature change from the SSP. The term $\delta_{i,t}$ is the temperature-induced increment (or decrement) to growth due to temperature changes from country-specific recent historical averages. Specifically, it evolves according to $\delta_{i,t} = f(X_{i,t+}) - f(\overline{X}_i)$, where $X_{i,t+}$ is projected temperature beyond 2010 and $\overline{X}_i$ is country-specific average temperature from 1980 to 2010. The function $f(X)$ is the estimated temperature function (polynomial or spline) of temperature from Eq. (1). Those estimated functions for all 800 specifications are shown in appendix Fig. A5. These temperature functions are assumed to be constant to 2100, and, thus, do not admit adaptation to temperature changes.[42] Temperature deviations are estimated by assuming a linear increase from the historical average to country-specific temperature projections in 2100 from the RCP8.5.

BHM demonstrate the uncertainty of GDP climate impacts by bootstrapping the estimation of the growth-temperature relationship. In their preferred specification they find that approximately 30% of bootstrapped simulations yielded positive global GDP gains from projected warming in 2100 even though the central estimate was a 23% loss. This demonstrates the substantial *sampling* uncertainty over future climate change damages attributable to uncertainty over parameters. We also estimate the magnitude of sampling uncertainty, but we uniquely relate it to *model* uncertainty, i.e., that uncertainty attributable to ambiguity about the correct model. As reported in the next section, model uncertainty is shown to rival even the substantial sampling

[41] This is true in forecast and backcast CV due to the inability to estimate out of sample fixed effects. In K-fold CV, we can estimate the fixed effects directly, so in that case we do not demean by $\widehat{\alpha}_i$ and $\widehat{\lambda}_{r,t}$ *ex ante*, and those terms are then included in the prediction calculation in equation (2).

[42] BHM estimate their preferred temperature function separately for the first half and for the second half of the data. They find no statistical difference in the mean temperature response across these subsamples, and, therefore, conclude there is no evidence of adaptation that should be incorporated into projections of the future effects of climate change. We similarly compare estimates for the first half and for the second half of our data and for all models that are included in an MCS. Fig. A3 reports the confidence region for the difference in parameter estimates derived from the early subsample and the late subsample for 1000 bootstrap samples from each early and late subsamples of the data. Like BHM, we find no evidence of parameter heterogeneity across these time periods.

uncertainty BHM estimate, unless one focuses solely on models that relate non-linear temperature to GDP levels.

4. Results

This section presents the results of the cross validation exercise, comparing the cross-validated root-mean-square errors (CV RMSEs) across all 800 models. We then illustrate the estimated relationships between GDP and temperature across all models favored in the previous literature and those favored by cross validation. The estimated GDP impact in 2100 under each model specification is illustrated for the benchmark scenario of unmitigated warming (i.e., RCP8.5). Finally, impact heterogeneity across rich and poor countries and across agricultural and non-agricultural production is explored.

4.1. Model cross-validation

Employing cross validation by three distinct methods across 800 models results in 2400 estimates of RMSE. The RMSE for each model i is calculated as $RMSE_i = \sqrt{n^{-1}\sum_t e_{i,t}^2} = \sqrt{n^{-1}\sum_t (Y_{i,t} - \widehat{Y}_{i,t})^2}$, using the actual and predicted values from the test set, $Y_{i,t}$ and $\widehat{Y}_{i,t}$, respectively.

4.1.1. Forecast CV

Forecast cross validation reveals that dozens of alternative models are characterized by similar predictive ability as summarized by RMSE. These results are reported in Fig. 1, which depicts for each of 800 models the estimated RMSE and the associated 95% confidence interval. The figure is split into two panels; the top panel shows models with year fixed effects, and the bottom panel shows the models with region-year fixed effects. The area below the depicted RMSEs indicates by gray cell color the characteristics of the respective model specifications. For instance, reading from left to right, the first half of models in each panel are characterized by GDP level effects, and the latter half are those estimating growth effects, as indicated by the cell colors in the bottom two rows of the specification chart. For legibility, the y-axis in the figure is truncated at 0.12 (twice the sample standard deviation) because some models with high order time trends exhibit very high RMSEs.

The green dots in the bottom panel indicate RMSEs for those models with region-year fixed effects that appear in the MCS. Blue dots in the top panel denote RMSEs of the year fixed effect models appearing in the MCS. The red dot indicates the RMSE of the BHM model (which is not in the MCS), and black dots depict RMSEs of all remaining models.[43] The figure demonstrates that 63% of models are characterized by RMSEs above the sample standard deviation of 0.0603, suggesting a simple prediction equal to mean GDP growth is more a fccurate than many of these models. As shown later in this section, common predictive ability yields a large set of models of superior predictive ability that is statistically indistinguishable across models within the set. This, combined with sensitivity of projected GDP losses from warming, implies tremendous model uncertainty that is further considered later in this section.

Prediction errors are similar across level and growth models and models that include year or region-year fixed effects if county time trends are excluded. Models that include time trends, however, perform poorly in cross validation relative to models that exclude them. This is particularly true for models that include region-year fixed effects, and it is indicative of over-fitting.[44] In fact, forecast CV prediction errors are minimized by excluding time trends irrespective of other modeling assumptions, as indicated by the 9th row of the specification chart in Fig. 1 (labelled "Linear Trend"). RMSE increases as higher-order polynomial trends are included. Among models that exclude trends, RMSE varies from 0.0497 to 0.0535. Models with trends have uniformly higher RMSEs. The RMSE of the BHM model is 0.0612. The inferior out-of-sample fit among models with time trends (as in BHM) is notable because BHM indicate their preference for quadratic time trends is partly due to *in-sample* prediction accuracy. As we show in section 4.1.3, models that include parametric trends are excluded from model confidence sets, i.e., they are estimated to be statistically inferior to models that exclude such trends, conditional on the forecast procedure. Hence, forecast CV results illustrate that in-sample predictive accuracy can give a misleading view of out-of-sample validity.

Given the interest of a subset of the climate econometrics literature in understanding the temperature sensitivity of GDP and GDP growth, it is striking that among models that exclude country time trends, RMSE is largely invariant to the exclusion of temperature, or to inclusion of any of the temperature functions we model. For instance, among models without trends, RMSE varies by less than 1 percent across temperature specifications holding other model choices constant. Similarly, model performance is also insensitive to the inclusion or exclusion of temperature lags. The invariance of RMSE to temperature specification is, perhaps, unsurprising given the multiple factors determining GDP and its growth and the relatively small share of variation explained by temperature.

4.1.2. Backcast and K-fold CV

Results of backcast and K-fold CV are depicted in Figs. A7 and A8, respectively. Backcast and forecast CV are methodologically similar in that both require extrapolation across five consecutive years of the test data. Results across CV methods are qualita-

[43] Models that cannot be seen because their RMSEs exceed the point of truncation of the y-axis do not appear in any MCS.

[44] RMSEs for these models exceed the top range depicted in Fig. 1 due to the truncation of the y-axis.

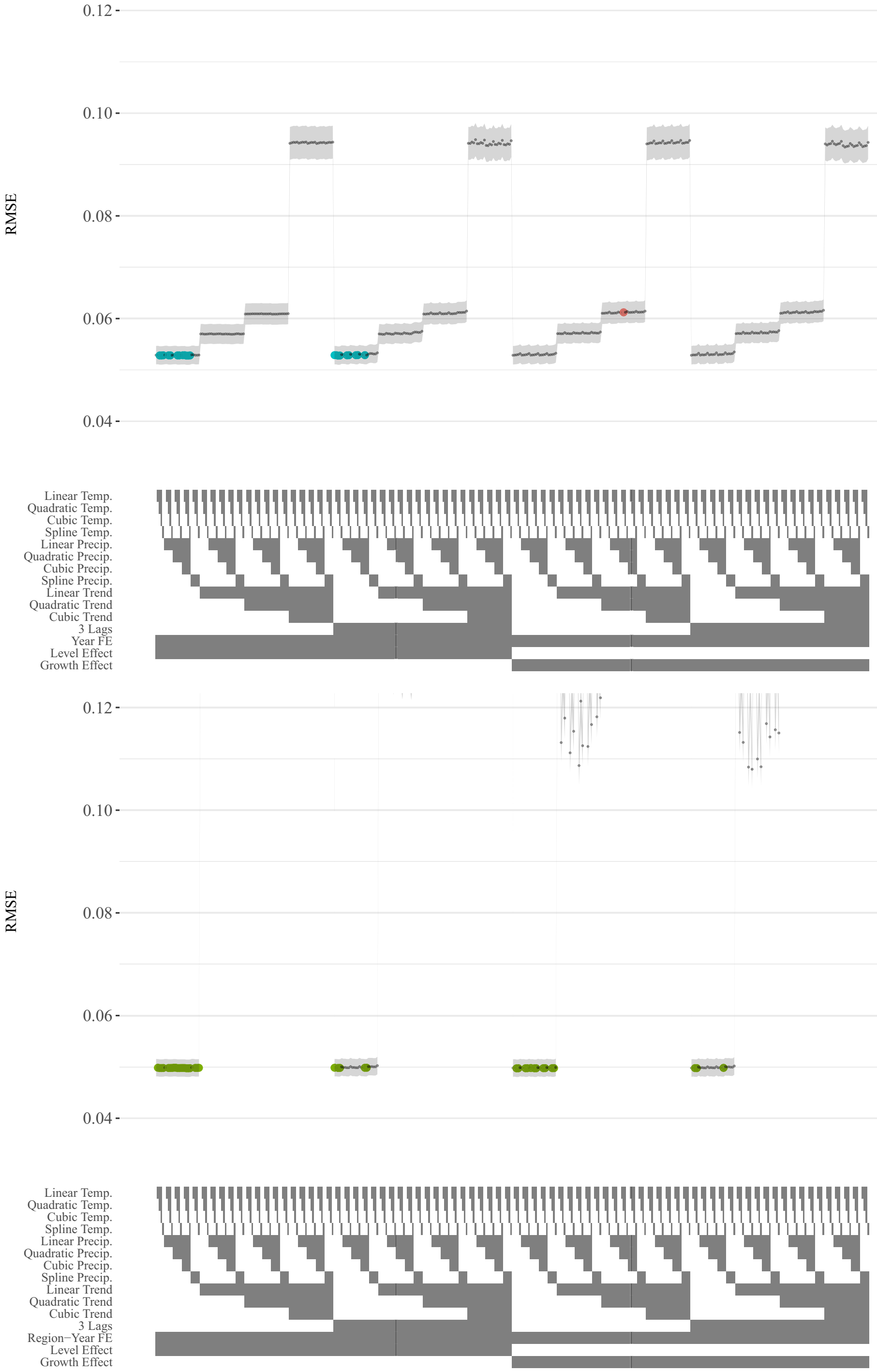

Fig. 1. RMSE under Forecast CV (20 Year Estimation Window), by Specification. Year FE models (top) and Region-Year FE models (bottom). MCS models in blue and green in top and bottom figures, respectively. Notes: This figure depicts RMSE under forecast CV for 800 models. Green dots indicate RMSE for region-year FE models appearing in the $T_{\max,\mathcal{M}}$ MCS under this CV approach. Blue dots are year FE models in the MCS. The red dot is the BHM specification, which does not appear in the MCS. Black dots are the remaining models. Gray areas represent ± 1.96 standard errors of the RMSE. For legibility, the y-axis is truncated at 0.12 (twice the sample standard deviation); all models obscured by this truncation have very high RMSEs and accordingly do not appear in any MCS.

tively similar, favoring models that exclude flexible parametric trends. RMSE increases in backcast CV as higher-order polynomial trends are introduced. Regardless of other modeling assumptions, RMSE is minimized by excluding country trends. Likewise, regardless of other modeling assumptions, prediction error is least among models that include region-year fixed effects. Among those models that exclude country-specific trends, RMSE varies from 0.0488 to 0.0515, or 81–85% of the sample standard deviation of GDP growth (0.0603).

Like forecast CV, model performance assessed using backcast and K-fold CV is largely invariant to how temperature is modeled or whether it is excluded. Holding other model specifications constant, RMSE varies across temperature functions by less than 1%. Conditional on other model choices, RMSE is also insensitive to whether GDP growth or level effects are modeled, and whether temperature lags are included or not.

Prediction errors across models in K-fold CV vary little, but best-performing models are similar to those identified by forecast and backcast CV. Irrespective of other model characteristics, RMSE is minimized among models that include region-year fixed effects and exclude parametric trends. It is insensitive to temperature or precipitation function specifications. Holding other model characteristics constant, the RMSEs vary by less than 0.3% across temperature specifications. Among models that include year fixed effects and not region-year fixed effects, RMSE declines as higher-order polynomial trends are added until a cubic trend, which yields the largest RMSE irrespective of other model characteristics.

4.1.3. Model confidence sets

We estimate model confidence sets comprised of 60, 53, and 32 models for the forecast, backcast, and K-fold CV, respectively. The MCS is analogous to a confidence interval on a parameter estimate in that it is assured to contain the best-performing model at a given confidence level. Like the confidence interval on a parameter estimate, the greater size of the MCS reflects greater uncertainty, i.e., the limits of information in the data from which to identify the best model. The fact that the size of the MCS reflects the information content of available data makes the MCS procedure attractive relative to other tests for superior predictive ability (Hansen et al. 2011).

The prediction errors of models retained in the MCS are indicated by blue or green dots in Fig. 1, A7, and A8. Blue dots in Fig. 1 and A7 indicate the RMSEs of year fixed effects models of statistically superior performance. Green dots in the figures are analogous indicators for models composed of region-year fixed effects. Because only one MCS is estimated for K-fold CV, the RMSEs of models contained in the MCS are indicated by green dots. (The red dot in each figure indicates the BHM model.)

The MCSs identified by forecast CV RMSEs are exclusively comprised of models that exclude the parametric time trends preferred by BHM. The MCSs, however, do not discern among temperature functional forms or growth and level effects. The forecast MCSs contain models including all temperature and precipitation specifications, as well as models that specify growth and GDP levels relationships. The MCSs selected by backcast CV RMSEs similarly exclude any parametric time trends and include models specified by each temperature and precipitation function, as well as growth and levels models. The model preferred by BHM is excluded from all MCSs. Models similar to DJO, which we term "DJO∗" and "DJO∗+Quad. Temp.", are included in all MCSs, as are some models that include temperature lags.[45]

4.1.4. Summary of CV results

We conclude the following from CV and MCS analyses:

1. Dozens of alternative models exhibit comparable predictive ability in cross validation. This leads to large sets of models characterized by statistically indistinguishable predictive ability. The range of RMSEs for models included in the MCSs is 0.0497–0.0566.
2. GDP growth and levels models exhibit similar predictive ability, and we cannot identify with 95% confidence whether levels or growth models have superior predictive performance. Neither can we determine with 95% confidence whether the most predictive models include temperature lags or not.
3. We cannot preclude at the 95% confidence level that the most predictive model excludes temperature or adopts any of the temperature functions we considered. The invariance of RMSE to temperature specification is, perhaps, unsurprising given the multiple factors determining GDP and its growth and the relatively small share of variation explained by temperature.[46] Model predictive ability is also invariant to the specification of the function relating precipitation to GDP or GDP growth.
4. Models that include parametric country-specific time trends, as in BHM, are excluded from model confidence sets derived from forecast or backcast CV, and the model preferred by BHM is excluded from all MCSs.

[45] DJO∗ includes a linear temperature function and region-year fixed effects. This adaptation of DJO excludes interactions with indicators of countries' below-median-income status in order to nest the model within the dimensions of the model space we defined. DJO∗+Quad replaces the linear temperature function with a quadratic.

[46] A linear function of temperature explains only 0.2% and 0.03% of sample variation in GDP and GDP growth, respectively, conditional on region-year and country fixed effects. Even allowing for the non-linear effects of temperature, a flexible four-knot cubic spline of temperature explains only 0.9% and 0.4% of GDP and GDP growth variation, respectively.

4.2. Marginal effects of temperature

4.2.1. Overall marginal effects

Previous studies have identified statistically significant temperature effects on GDP or GDP growth within the specific models chosen by the authors. Such findings account for sampling uncertainty, but they do not account for uncertainty over the correct specification of the model relating economic aggregates to temperature. Given the number of models comprising our estimated model confidence sets, we assess the significance of marginal temperature effects across those models comprising the union of MCSs. We, thus, systematically assess the robustness of the findings in earlier studies rather than assessing the significance of marginal temperature effects in select alternative models. We proceed by estimating 1000 bootstrap samples for each model in the union of the MCSs, generating a distribution of estimated marginal temperature effects for each model.[47] These distributions are aggregated across models to yield a distribution of marginal temperature effects reflecting sample and model uncertainty. From this distribution, confidence regions are reported for annual temperatures on the support of the historical record.

Fig. 2 reports the mean temperature marginal effects separately for growth and levels models, along with confidence regions. The mean marginal effect is shown in black. The green-shaded region reflects a 50% confidence region, and each successive band of shading reflects an expansion of the confidence level. A 95% confidence region, for instance, is given by the red-shaded bands and the area between them.

For growth effects models, the top panel of Fig. 2 shows that the mean marginal effect is positive at temperatures below 13.4 °C, and negative thereafter. The monotonic and approximately linear decline in the marginal temperature effect characterizes a quadratic temperature function with mean peak at 13.4 °C. However, the effect of historical temperature variation on GDP growth is very imprecisely estimated and includes zero for all reported confidence regions and across all temperatures of the support. This is true at even the 50% significance level, and is shown by the confidence regions spanning positive and negative marginal effects in the top panel of Fig. 2.

Turning to the bottom panel of Fig. 2, the mean marginal effect of temperature on GDP levels declines nearly monotonically from approximately a 1.5% gain to GDP per degree of warming at the coldest annual temperatures observed during our study period, to an approximately 1.9% marginal decline in GDP at 30 °C. As above, the marginal effect exhibits a roughly linear decline, characteristic of a quadratic temperature function with a mean peak at 11.8 °C, where the marginal effect crosses the axis. The marginal effect of temperature on GDP levels is more precisely estimated than the effect on GDP growth, yet there is still a wide range of temperatures for which statistically significant temperature effects are not identified at conventional levels of statistical significance.

These results suggest there is far less certainty about the magnitude of marginal temperature effects on aggregate GDP than some prior studies imply. Whereas other studies, such as BHM and DJO, have identified statistically significant growth effects when accounting for sampling uncertainty alone, we have shown that the contribution of model uncertainty renders those effects indistinguishable from zero at conventional levels of certainty, and points to effects on GDP levels as being more confidently detected. The substantial model uncertainty is a consequence of the similar out-of-sample performance of alternative models that, nevertheless, imply disparate marginal effects of temperature on GDP growth or levels.

4.2.2. Comparing levels and growth models

We further disaggregate the marginal effects shown in Fig. 2 into contemporaneous and lagged effects for those models among the MCS that admit such lags in order to assess whether temperature exhibits an effect on growth, which persists, or a transitory effect on GDP levels. Fig. 3 depicts for growth models the total marginal temperature effect, the contemporaneous marginal temperature effect, and the lagged temperature marginal effect. The contemporaneous marginal effect (β_0) and the lagged marginal effect ($\beta_1 + \beta_2 + \beta_3 + \beta_4$) are shown to be of approximately equal magnitude but opposite sign, consistent with transitory, level effects of temperature on GDP, rather than growth effects. This is consistent with the sign reversal observed across contemporaneous and first-lag temperature effects in DJO and BHM. We do not identify statistically significant marginal effects of three temperature lags across the temperature support, as shown in the bottom panel of the figure. The contemporaneous effect of temperature is shown to be statistically significant at the 10% level for high and low temperatures, as shown in the middle panel of the figure. The composite of lagged and contemporaneous temperature effects is statistically indistinguishable from zero across the temperature support as shown in the top panel of the figure.

4.2.3. Implied optimal temperatures

Uncertainty over the magnitude of temperature marginal effects yields substantial uncertainty as to the optimal temperature for production or economic growth using these aggregate economic measures. Model uncertainty in the optimal temperature is reflected in Fig. 4, which depicts a histogram of the GDP-maximizing temperature implied by those models included in the union of MCSs and characterized by an interior maximum on the temperature support. The shading and color of the histogram indicate which fraction of models predicting each temperature optimum is characterized by growth or level effects and the inclusion or exclusion of temperature and precipitation lags.

[47] The bootstrap was implemented by sampling coefficients from the cluster-robust variance-covariance matrix (clustered by country and year).

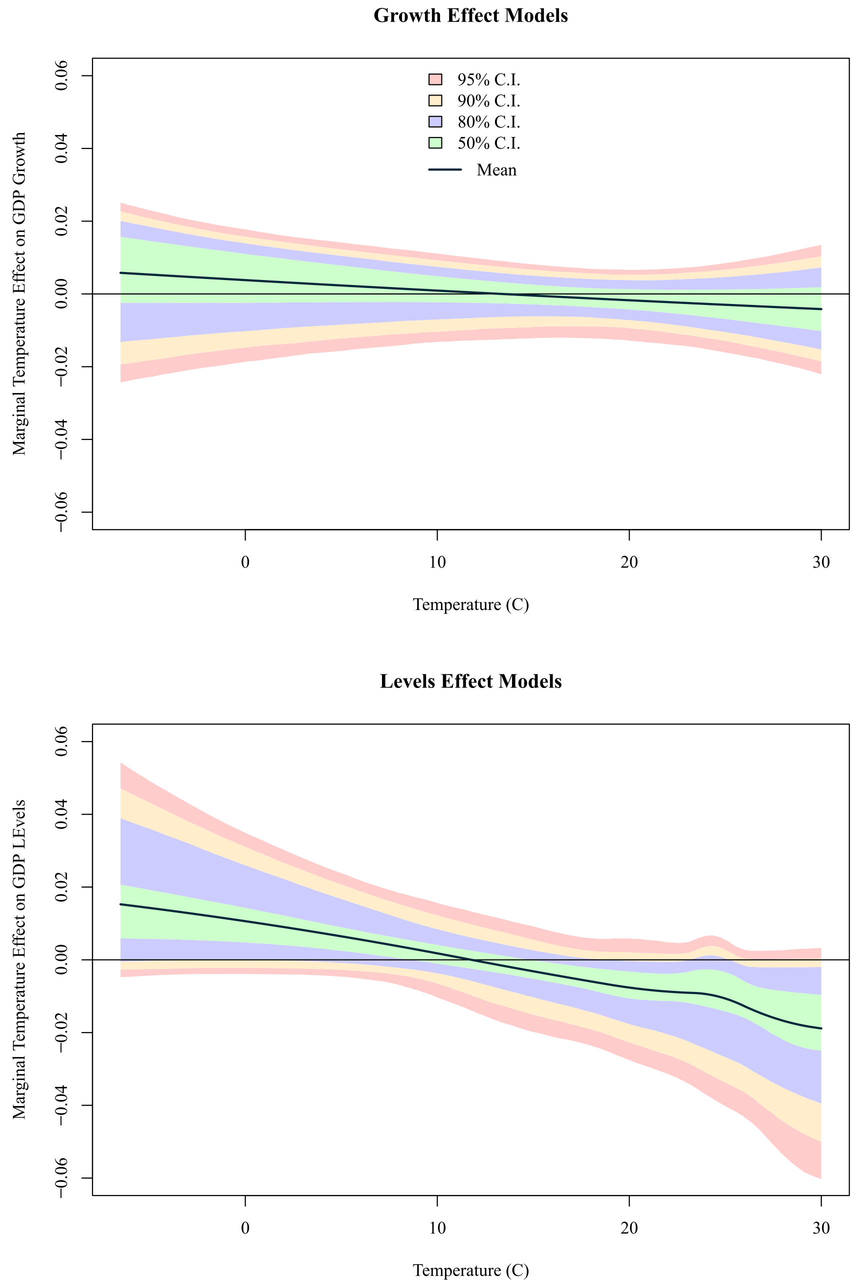

Fig. 2. Confidence regions of temperature marginal effects on $\Delta \ln(GDP)$ (top) and $\ln(GDP)$ (bottom). Notes: Plotted are 50, 80, 90, and 95% confidence regions of temperature marginal effects on GDP growth (top panel) and GDP levels (bottom panel). Confidence regions are determined by bootstrapped sampling 1000 times for each model specification that appears in the union of model confidence sets.

As Fig. 4 shows, model uncertainty admits a wide range of GDP-maximizing temperature even among models determined to be superior in the model selection procedure. While the central tendency of the estimates (mean and median) is in the range of 12–13.2 °C, these models imply a range of optimal temperatures of about 10 °C to greater than 20 °C. GDP growth models that exclude lags exhibit the narrowest distribution of GDP-maximizing temperature: 15.2–15.8 °C. GDP growth models that include lags imply considerably higher optimal temperatures than other models, ranging from 19.4 to 20.4 °C. The optimal temperatures under levels models without lags range from 10.8 °C to 13.8 °C. The range of optimal temperatures is widest for levels models that include lags: 10–15.4 °C.

The differences across models in estimated temperature optima imply considerable variation in projected climate change impacts because the bulk of global economic production typically experiences temperatures near the mean and median of this distribution: 58% of global GDP in 2010 derived from countries with temperatures between 10 °C and 15 °C. For instance,

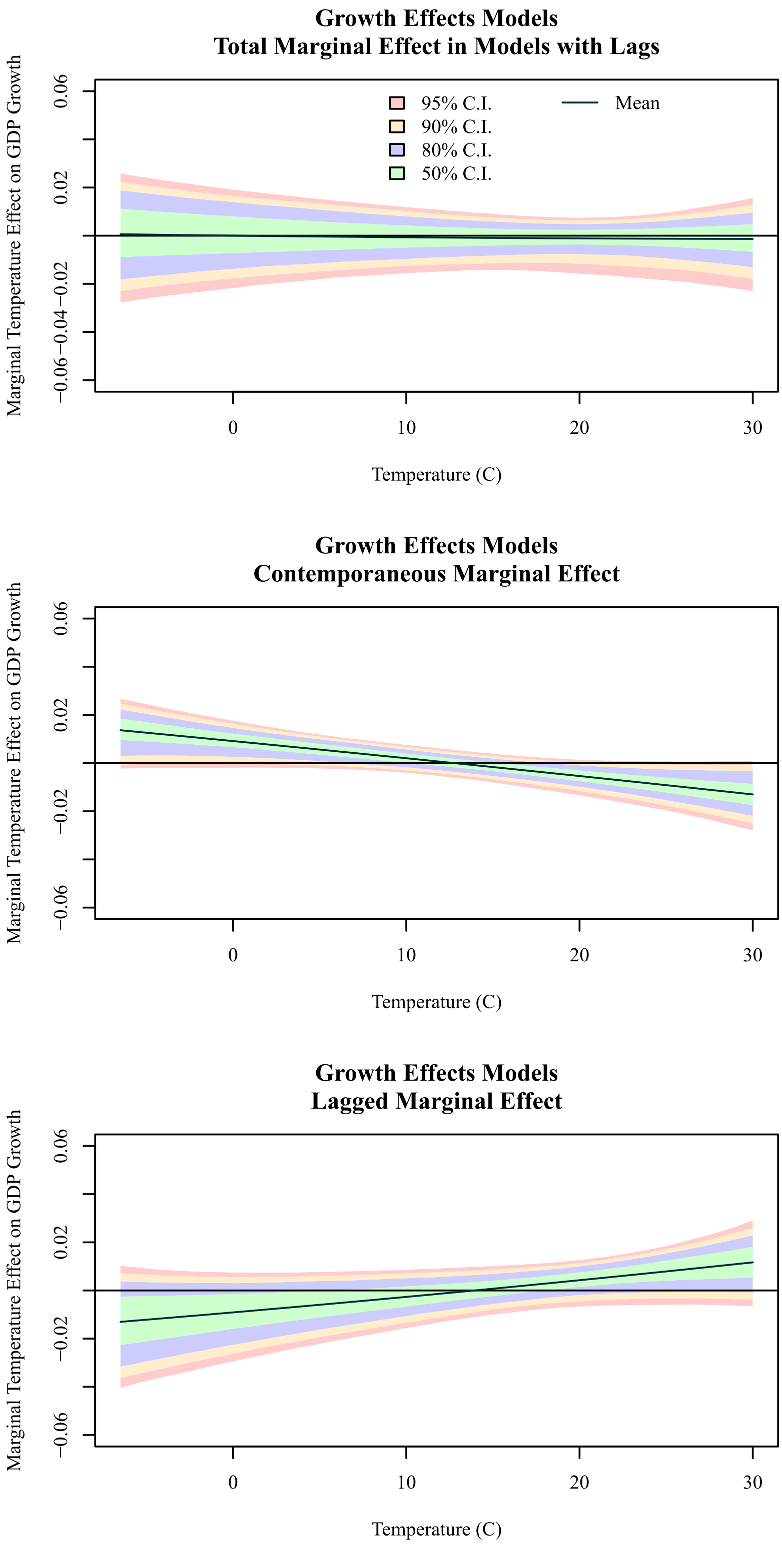

Fig. 3. Confidence regions of total, contemporaneous, and lagged temperature marginal effects, among MCS growth models with lags. Notes: Plotted are means and 50, 80, 90, and 95% confidence regions of total (top), contemporaneous (middle), and lagged (bottom) temperature marginal effects on GDP growth. Confidence regions are determined by bootstrapped sampling 1000 times for each model specification that appears in the union of model confidence sets.

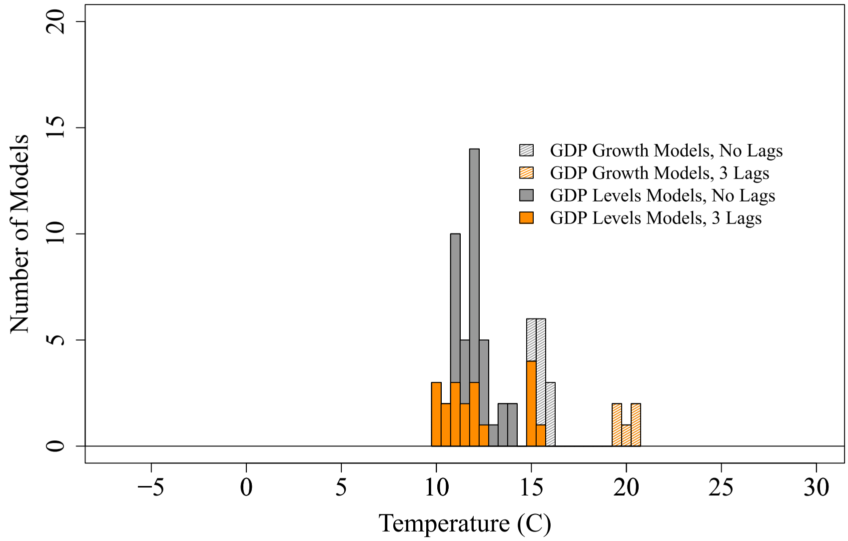

Fig. 4. Histogram of GDP-maximizing Temperatures Among Models in Any MCS. Notes: This is a histogram of GDP-maximizing temperatures for models in the union of MCSs that are characterized by an interior maximum over the support of temperature. This excludes all linear models, but it also excludes some growth models with lags that are generally downward sloping over the temperature support; see Fig. A6.

average temperatures in the three largest economies in 2010—the United States, Japan, and China—were about 14 °C, above the temperature peak estimated by BHM of 13 °C. Hence, these countries would be expected to suffer GDP losses from warming. But these economies operate below the estimated "DJO∗+Quad" temperature peak of 15.4 °C, implying benefits from 2 °C of warming. These three economies collectively accounted for over 40% of global GDP in 2010.

Sampling uncertainty is more substantial. The 95% confidence interval of the temperature optimum in the "DJO∗+Quad" model, for instance, is 7.8–30 °C. We further assess the combined sample and model uncertainty by bootstrapping 1000 draws for each model in the union of MCS.[48] The 95% quantile interval of temperature optima extends beyond the support of the historical record of annual average temperature.

4.3. Estimated GDP impact of climate change

As described in section 3.5, the parameter estimates from the 800 model we estimate are used to project economic impacts of unmitigated warming. As in BHM, we project the impact of expected warming on global GDP by 2100 using the RCP8.5 climate projection as a benchmark of unmitigated climate change and SSP5 for projections of baseline GDP and population growth.[49] In addition to examining the variation in GDP impacts due to model uncertainty, we evaluate sampling uncertainty using a bootstrap procedure for the subset of models appearing in any MCS and including non-linear temperature functions. We are, thus, able to relate sampling uncertainty to model uncertainty and assess the variance in projected GDP losses due to both.[50]

Modeling Uncertainty. Table 2 shows the estimated damages for each model that incorporates a quadratic of precipitation and excludes temperature lags. The top panel of the table reports projections generated by growth models; the bottom panel shows projections of levels models. Growth models project a wide range of GDP impacts in 2100, whereas levels models project a considerably narrower range of smaller impacts. The variation in estimated GDP impacts across models is due primarily to the importance of the GDP-maximizing temperature level relative to the location of the world's major economies. Small shifts in the GDP-maximizing temperature can change whether GDP of a few major economies is estimated to benefit from or be harmed

[48] The bootstrap procedure is described in section 4.3 under *Sampling and Modeling Uncertainty*. We report the 95% quantile range rather than standard errors because the distribution of the peak is non-standard. For example, the peak of a quadratic, ($\frac{a}{-2b}$), has a fat-tailed and asymmetric distribution as the ratio of two normally distributed random variables.

[49] We abstract from modeling weather variability. Instead, following BHM, we project future temperatures as a deterministic function of the climate warming projected by RCP8.5. In particular, temperature is assumed to evolve according to a constant annual increment to country-specific temperature means from 1980 to 2010, as explained in Section 3.4. As can be shown, weather variability contributes to additional variability of future GDP impacts of climate change among non-linear growth models. (We are grateful to an anonymous reviewer for contributing this insight.) For growth models with non-linear temperature functions, weather shocks interact with the effects of climate change, and this interaction effect has permanent consequences. For linear models, weather variability does not interact with climate effects, and, thus, does not contribute to the variance of the climate effect. For models of GDP level effects, the consequences of weather variability are not permanent.

[50] Following BHM, we also abstract away from uncertainty in the baseline GDP projection. This likely understates the uncertainty in GDP impacts, which in turn depend on uncertainty in economic growth. Indeed, Müller et al. (2019) demonstrate considerable uncertainty in baseline economic growth. We leave an exploration of this to future research.

Table 2
Estimated percentage effects of unmitigated warming on GDP by 2100 by model including quadratic of precipitation, No Lags.

	Temperature Functional Form			
Climate GDP Impact in 2100 (%)	Linear	Quadratic	Cubic	Spline
GDP Growth Effect				
Year FEs; No time trend	-1.02	-13.25	-16.39	-7.75
Year FEs; 1 degree time trend	-35.38	-46.53	-38.36	-37.07
Year FEs; 2 degree time trend	-37.21	**-49.44**	-49.59	-49.46
Year FEs; 3 degree time trend	-42.51	-54.26	-54.73	-55.86
Region-Year FEs; No time trend	22.27	5.63*	3.61*	32.55
Region-Year FEs; 1 degree time trend	-21.54	-32.90	-19.46	-21.10
Region-Year FEs; 2 degree time trend	2.33	-17.41	-6.26	-6.13
Region-Year FEs; 3 degree time trend	-6.21	-27.32	-19.20	-22.17
GDP Level Effect				
Year FEs; No time trend	-0.43	-1.71*	-1.63*	-2.17*
Year FEs; 1 degree time trend	-0.52	-1.83	-1.73	-2.23
Year FEs; 2 degree time trend	-0.63	-1.97	-1.88	-2.33
Year FEs; 3 degree time trend	-0.81	-2.13	-2.04	-2.51
Region-Year FEs; No time trend	-0.64*	-1.82*	-1.75*	-2.16*
Region-Year FEs; 1 degree time trend	-0.69	-1.90	-1.83	-2.22
Region-Year FEs; 2 degree time trend	-0.53	-1.80	-1.72	-2.04
Region-Year FEs; 3 degree time trend	-0.58	-1.82	-1.74	-2.07

Notes: * are models in the MCS under any of the three main CV approaches. BHM's chosen specification in red. GDP impact is estimated to be greater than in BHM because of a 4% reduction in sample size to accommodate three temperature lags that are included in some models.

by projected warming. Models that comprise the MCSs, denoted by (∗) in the table, exhibit a much narrower range of projected impacts.

GDP losses are highly sensitive to modest changes in modeling assumptions of growth models, as shown in the top panel of Table 2. Models that include year fixed effects all imply GDP losses by 2100. In contrast, models that include region-year fixed effects but exclude time trends, i.e., those that minimize RMSE, all imply GDP gains by 2100. The BHM model predicts GDP losses of 49%, greater than most models reported in the table.[51] If the BHM specification were more saturated with fixed effects, the associated loss would be 17% rather than 49%. If the BHM specification used the DJO controls for unobservable trends (i.e.,

[51] In order to accommodate three temperature lags, our sample is necessarily reduced in size by 4%. Thus, whereas BHM estimate a 23% decline in GDP by 2100 across their sample, we estimate a 49% decline from their same model estimated on the marginally reduced sample. This is emblematic of the sampling uncertainty that characterizes this econometric approach.

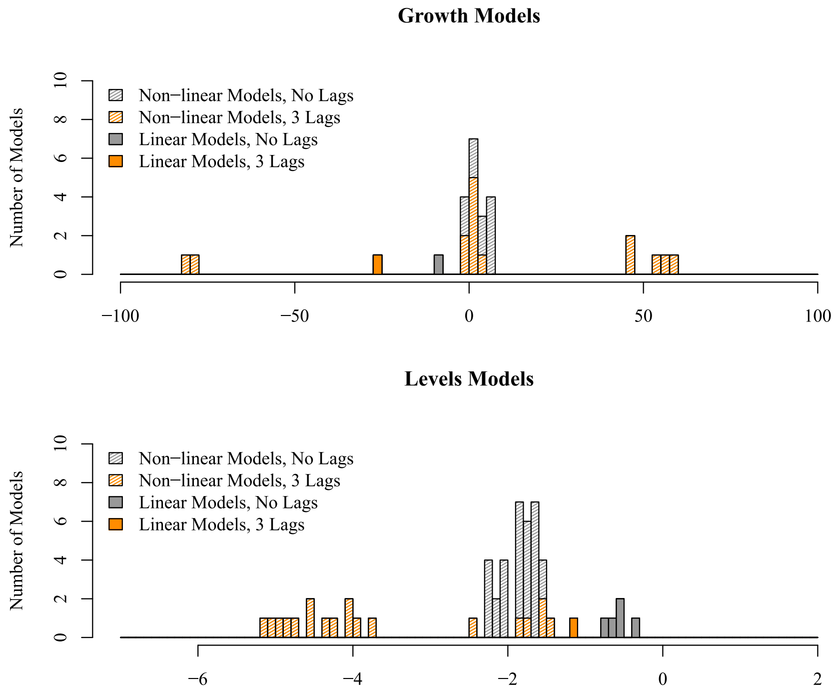

Fig. 5. 2100 GDP Damage Estimates in Unmitigated Warming Scenario for GDP Growth Models (Top Panel) and GDP Levels Models (Bottom Panel) in Any MCS. Notes: The stacked bars indicate models featuring linear versus non-linear temperature functions, and models with lags versus those without. As seen in the bottom panel, the multi-modal nature of the distribution is driven by the linearity of the temperature functional form and the inclusion of lags.

region-year effects only) it would yield a modest GDP gain rather than a substantial GDP loss.[52] The DJO-adapted model (DJO*) implies GDP gains of 22% by 2100, while the DJO-adapted model that includes a temperature quadratic (DJO* +Quad. Temp.) implies gains of 6%.

In contrast to the growth models, nearly all levels-effect models (without lags) imply a narrow range of GDP losses equal to 2.5% or less, as reported in the bottom panel of Table 2. None of the 64 levels-effect models shown in Table 2 (and only 20 of the 320 levels models that include temperature functions) project GDP gains. Across model dimensions shown in the table, projected GDP losses are least among models that specify a linear temperature effect, and they are greatest among models the specify temperature splines. A GDP level model that adopts the other model characteristics of BHM projects GDP losses of 1.9% in 2100.

Fig. 5 shows the distribution of projected GDP impacts for all models included in the union of MCSs that specify a temperature function. It plots the histogram of percentage changes in GDP by 2100 from 27 GDP growth models (top panel) and 55 GDP levels models (bottom panel, note the different scale of the x-axis). In each panel, damages from models featuring linear and non-linear temperature functions are distinguished, as are those that include temperature lags and those that do not. Across all models in the union of MCSs specifying temperature functions, GDP impacts range from −80% to +59% without accounting for sampling uncertainty. This variation is attributed to growth models that include temperature lags, as shown in the top panel of Fig. 5.[53,54] In contrast, and consistent with the subset of results depicted in Table 2, GDP level models predict a narrow range of small GDP losses. The multi-modal distribution of projected losses among levels models is due to temperature specifications and temperature lags. Models that include non-linear temperature specifications and exclude lags typically project global GDP losses in the range of 1.5–2.5%, while those that include linear temperature functions project smaller losses. Non-linear GDP-levels models that include temperature lags project a wider range of economic impacts than other levels models, including losses of more than 5%.

Sampling and Modeling Uncertainty. Model and sampling uncertainty are separately compared for growth and levels mod-

[52] Table A1 is the analog to Table 2 for models estimated over the same sample as BHM. It demonstrates more starkly the sensitivity of projected GDP damages to model specification. For instance, growth models that exclude parametric time trends universally imply GDP gains from warming by 2100 whether year or region-year fixed effects are included. Modest changes to the specifications of BHM and DJO generate large changes in GDP impact predictions. Simply adding a cubic temperature term to BHM's specification reduces the estimated GDP impacts by half (−11% versus −23%). Removing BHM's country-specific time trends reverses the sign of the impacts (+12%), and using region-year fixed effects also reverses the sign (+10%). Starting with DJO's model and adding quadratic temperature reverses the sign of impacts (+41%).

[53] This is true despite many of them performing similarly well in cross validation. See Fig. 1 and A7.

[54] The BHM model is not depicted in the figure because it is excluded from all MCSs. However, in this sample, that model predicts GDP losses of 49%, greater than most models depicted in the figure and all such models that exclude temperature lags.

els by considering temperature parameter estimates from bootstrapped samples for each model that appears in an MCS and incorporates non-linear temperature functions.[55] Bootstrap temperature coefficients are drawn 1000 times from the clustered covariance matrix (clustered by country and year). Each set of temperature parameter estimates is used to project GDP losses in 2100, yielding a distribution of predicted global GDP impacts for each model in the union of model confidence sets.

These distributions are depicted in Fig. 6 for non-linear temperature models included in any MCS: the top panel shows distributions of 25 growth models, and the bottom panel shows distributions of 49 levels models (note different scales of horizontal axes). Each gray series depicts the distribution of GDP impacts for a distinct model in the union of the MCSs. Bolded blue curves in each panel depict the aggregate distributions of GDP impacts across all bootstrap samples and all non-linear temperature models in the union of the MCSs.

The top panel in Fig. 6 demonstrates considerable sampling uncertainty for models of GDP growth; individual model distributions range from 80 to 100% losses to greater than 500% gains for some models.[56] While the distribution of impacts across models has a peak below zero, it is centered at gains of +13% (median) to +47% (mean) due to the rightward skewness of the distribution. A small number of models, like that of BHM, are centered at losses of 50% or more, but most have substantial mass on GDP gains, resulting in positive medians and means of around +10 to +50%, due to a right skew. Whereas BHM acknowledge a 30% chance that future warming under RCP8.5 yields GDP gains, this likelihood is greater when accounting for model and sampling uncertainty. Fifty-seven percent of the mass of the pooled distribution is above zero. The 95 percentile range across these models and samples is −86% to +388%.

Sampling uncertainty is characterized by the average variance of GDP impact estimates holding models constant and varying samples. It can be related to the magnitude of model uncertainty, characterized by average variance of impact estimates as models vary and samples are held constant. The combined variance is the variance across all impact estimates from these models and is shown in the blue curve of the top panel. Model uncertainty, as measured by the average standard deviation of GDP impacts across these models, is equal to 83% of GDP. Sampling uncertainty is equal to 122% of GDP. Combined uncertainty is equal to 142% of GDP. The magnitude of combined uncertainty demonstrates the limited guidance to policymakers afforded by econometric specifications of the temperature-growth relationship.

The bottom panel in Fig. 6 shows less variation in predicted GDP impacts across bootstrap samples and across models in GDP levels. For most of the distributions depicted, GDP impacts are centered around −1% to −3%. The distributions centered around −5% are exclusively models with lags, as can be seen more clearly for the point estimates shown in Fig. 5. Ninety-five percent of the mass of the pooled distribution falls in the interval (−9.1%, +2.0%). This range is driven in large part by models admitting lags. The 95% range without those lag models is a tighter (−3.7%, −0.02%). The frequency of impacts worse than −3% GDP is about 30%, and there is only a 8% frequency of positive GDP impacts. Again, this 8% frequency is driven by the larger uncertainty associated with lag effect models; without these models, the frequency of positive GDP impacts is only 2.4%. Model uncertainty among these models is equal to 2.3% of GDP, whereas sampling uncertainty is 1.9%. Combined uncertainty is equal to 2.7% of GDP.

Among the models included in any MCS (including all models without regard to temperature functional form), the 95% quantile ranges of GDP impacts accounting for sample and model uncertainty are −84% to +359% for growth effects models and −8.5% to +1.8% for levels effects models. Even the narrowest MCS, based upon K-fold CV errors, admits considerable uncertainty in impacts, as it includes some GDP growth models; among those 32 models, the 95% confidence interval on GDP impacts ranges from −81% to +172%. This large range is, again, attributable to uncertainty about impacts in GDP growth models.

4.4. Temperature impacts by country income and sector

Rich vs. Poor Countries. BHM report that both poor and rich country GDP respond non-linearly to temperature.[57] This suggests climate change impacts on GDP are broader than those implied by DJO, who found negative impacts only on poor countries. We assess the robustness of these findings to alternate model specifications by estimating separately for rich and poor countries the marginal temperature responses from each model in the union of MCSs that include a temperature function. Means and confidence intervals of these marginal responses are reported in Fig. 7 for each temperature in the support of our data. They are reported separately for growth models (left panels) and levels models (right panels).

As shown in the top panels of the figure, the results do not identify a statistically and economically significant GDP growth or level response to temperature in rich countries at any annual average temperature observed in the historical record. Even a 50% confidence interval includes zero effect in growth models (see top-left panel). At temperatures typical of large economies, the 95% confidence interval extends from −1% to +2% of GDP at its narrowest point for levels models.

Among poor countries, the mean marginal *growth* effect of temperature exhibits a monotonic and approximately linear decline across the temperature support, implying a quadratic temperature function. This relationship is similar to the one characterizing the temperature marginal effect on aggregate GDP growth previously reported, suggesting poor country outcomes

[55] Theory and micro-foundations suggest a preference for non-linear temperature functions even though these models do not uniformly perform better in cross validation than models with linear temperature functions. By assessing uncertainty among only models with non-linear temperature functions, we constrain model uncertainty, which is nevertheless shown to be immense.

[56] The x-axis in the top panel is truncated at +500% GDP impacts, but the tail extends as high at +4700%.

[57] Deryugina and Hsiang (2014) also find significant temperature effects on U.S. GDP.

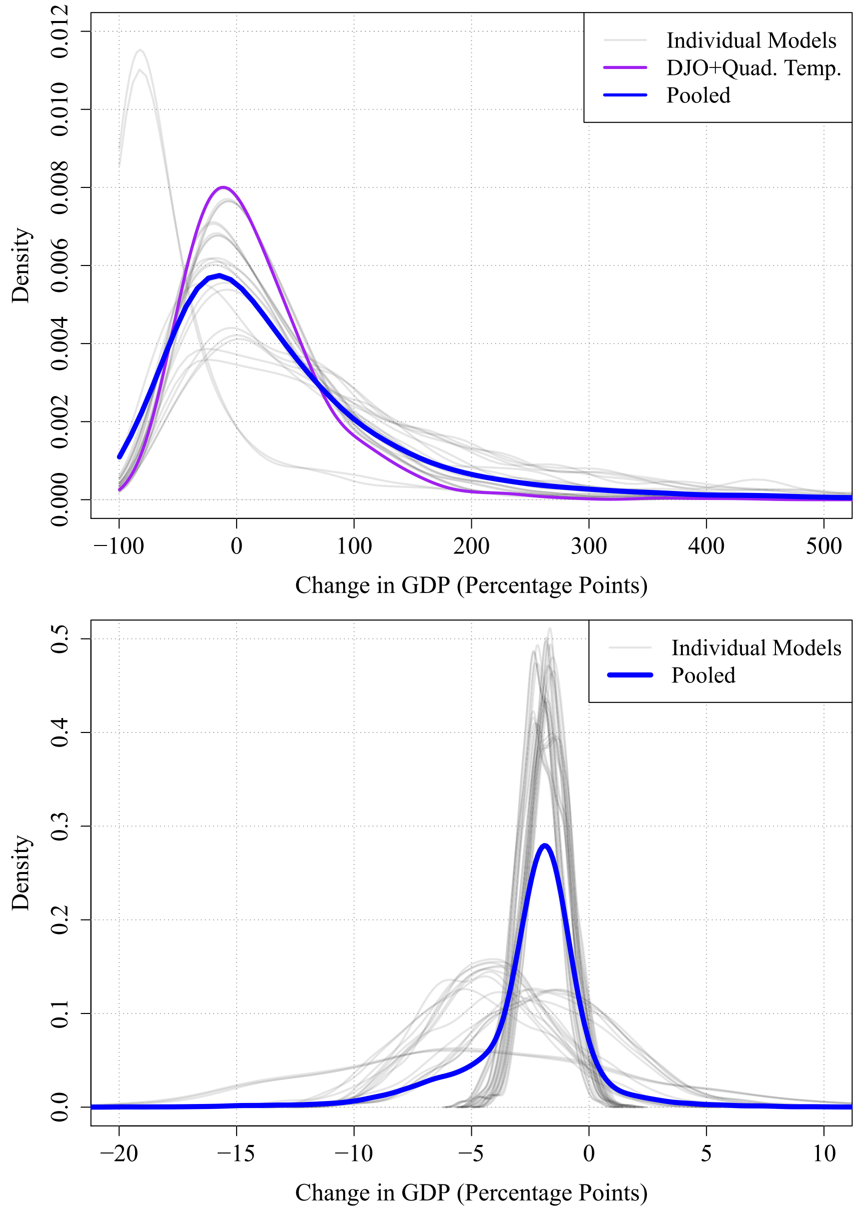

Fig. 6. GDP Impact Distributions for Unmitigated Warming in Non-Linear Models that Appear in Any $T_{\max,\mathcal{M}}$ MCS, GDP Growth Models (Top) and GDP Levels Models (Bottom). These are kernel densities of the bootstrapped GDP impacts for non-linear temperature models appearing in the $T_{\max,\mathcal{M}}$ MCS under any of the three primary CV approaches. There are 74 such models (25 growth and 49 levels). These distributions understate the full uncertainty in impacts because they assume no uncertainty in the other inputs to the projection (i.e., forecasts in baseline economic growth, population, and temperature). In the top panel, the x-axis is truncated at +500 percent GDP for legibility, obscuring less than 1.5 percent of the pooled mass. The levels models with central tendencies around −5 percent are lag models.

have an important influence on aggregate outcomes. However, the marginal temperature effect on poor country growth is imprecisely estimated, and even an 80% confidence interval includes zero for any temperature. The mean marginal GDP *level* effect among poor countries is negative across the temperature support, implying harms from warming at even the coldest annual temperatures observed over the historical study period (see bottom-right panel). The marginal effect declines nearly monotonically from −0.5% to −2.1% per degree of warming, and is more precisely estimated to be negative, particularly for temperatures above 18 °C.

Agriculture vs. Non-Agriculture Sectors. BHM and DJO also conclude that agricultural and non-agricultural production growth is affected by historical temperature shocks. However, taking both modeling and sampling uncertainty into account, we find more mixed results. This is shown in Fig. 8, which separately reports the confidence region of temperature marginal effects as in Fig. 7 for agricultural and non-agricultural production. The marginal growth effects of temperature are reported in the left panels, while marginal GDP level effects are reported in right panels. The top panels of the figure report non-agricultural

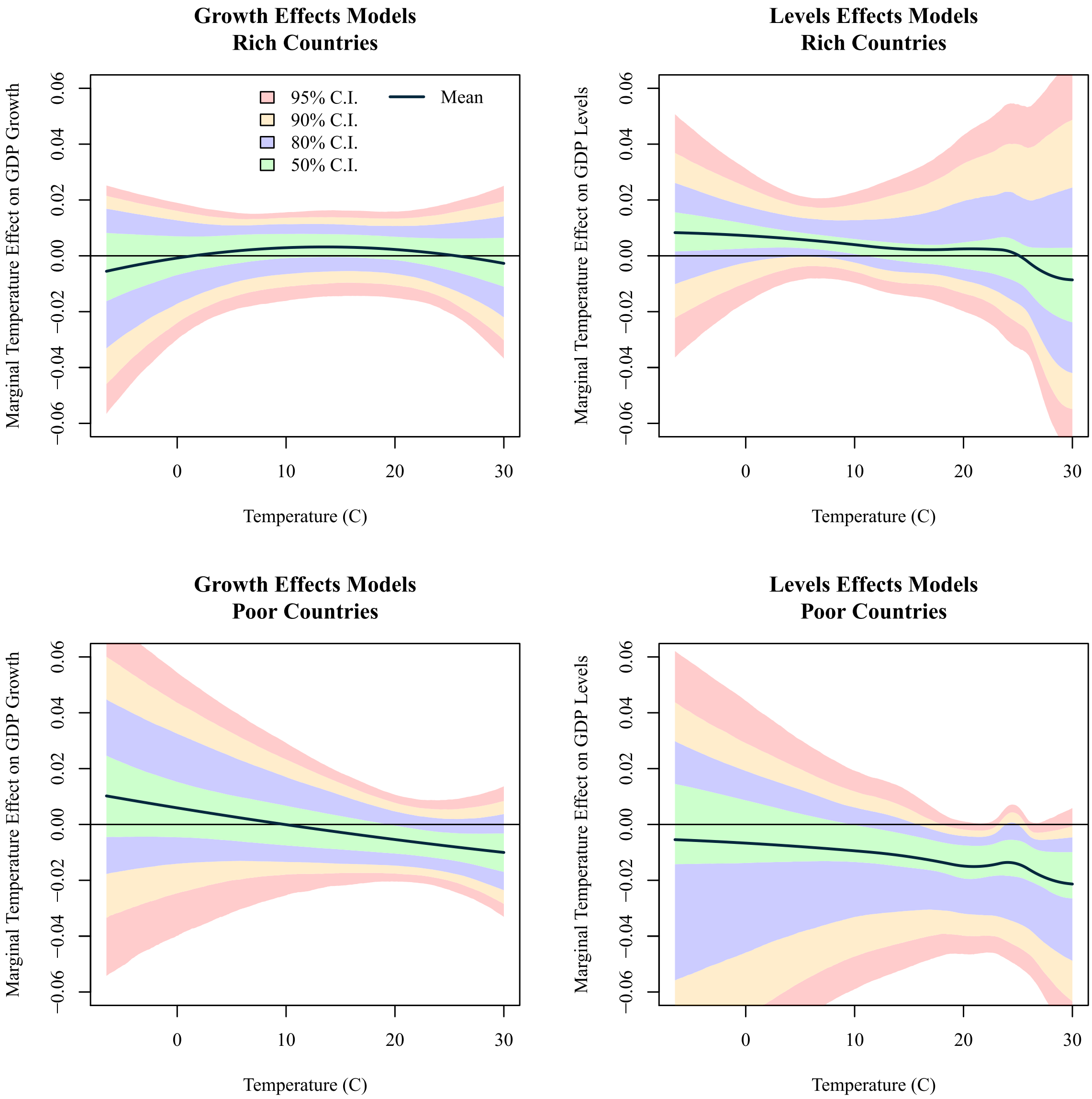

Fig. 7. Confidence regions of temperature marginal effects, Rich versus Poor Countries. Notes: Plotted are means and 50, 80, 90, and 95% confidence regions of temperature marginal effects on GDP growth (left column) and GDP levels (right column) for rich countries (top row) and poor countries (bottom row). Confidence regions are determined by bootstrapped sampling 1000 times for each model specification that appears in the union of model confidence sets.

production impacts, while the bottom panels report marginal effects for agricultural output.

The top panels show that the mean marginal temperature effect on both the growth and level of *non-agricultural* GDP is very imprecisely estimated and is only distinguishable from zero for a 50% confidence interval at very high temperatures in the levels effects model. Hence, we do not identify a statistically significant effect of temperature on non-agricultural production at conventional levels of statistical confidence.

In contrast, the effect of temperature on agricultural GDP shows a more consistent pattern. As show in the bottom panels of Fig. 8, the mean marginal temperature effect on the level and growth of agricultural output indicates a concave temperature function with a peak at about 10 °C. The marginal effect declines monotonically and approximately linearly across the temperature support in the growth effects model (bottom-left panel), but these effects are not statistically significant from zero across the temperature support at conventional confidence levels.

We find stronger evidence of substantial temperature impacts on agricultural GDP levels, as shown in the bottom-right panel of Fig. 8. The marginal mean effect is more steeply sloped than total GDP or any of the other disaggregated measures, and is statistically distinguishable from zero at the 20% significance level for temperatures above 15 °C.

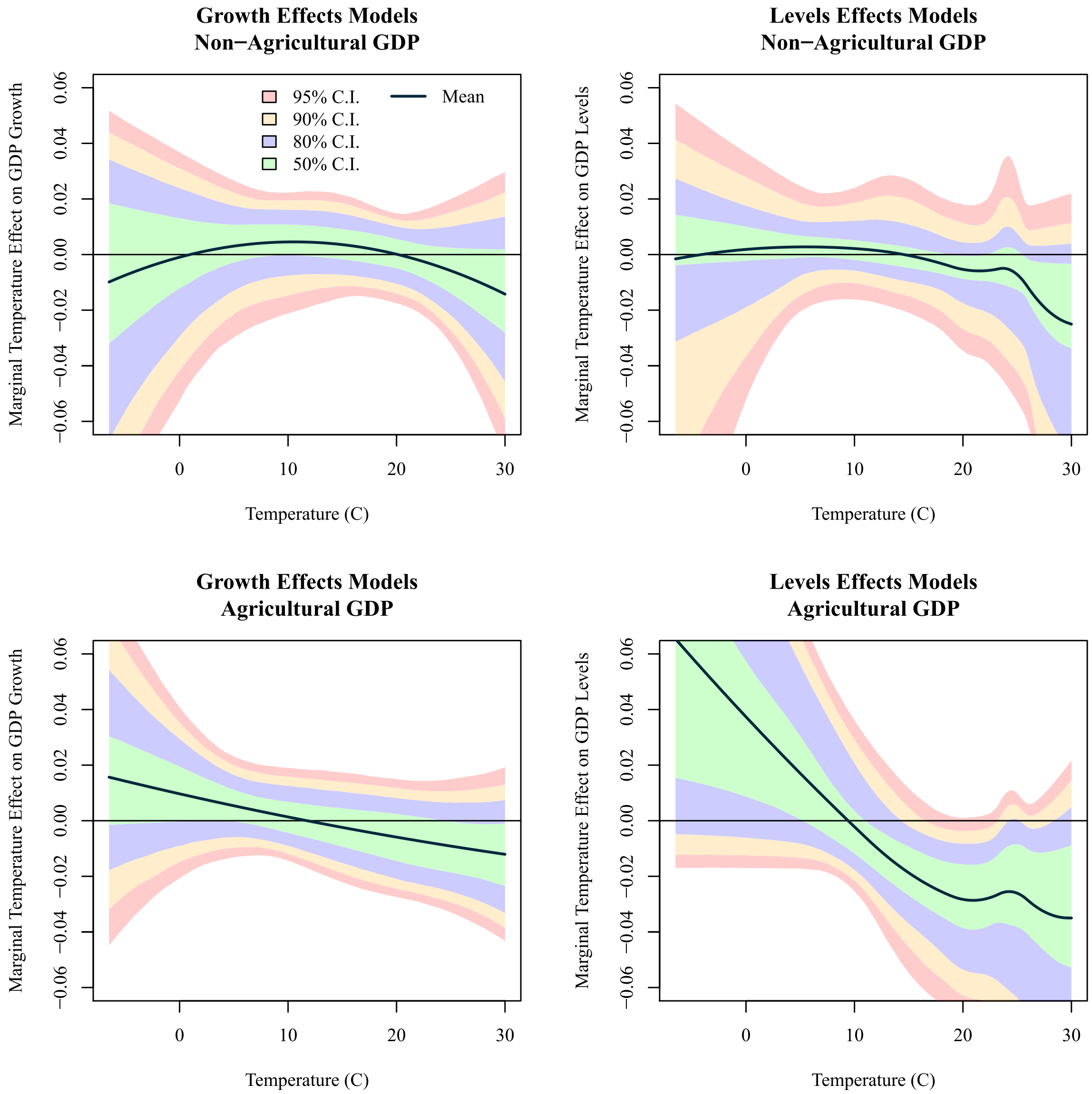

Fig. 8. Confidence regions of temperature marginal effects, Agricultural versus Non-Agricultural GDP. Notes: Plotted are means and 50, 80, 90, and 95% confidence regions of temperature marginal effects on GDP growth (left column) and GDP levels (right column) for non-agricultural GDP (top row) and agricultural GDP (bottom row). Confidence regions are determined by bootstrapped sampling 1000 times for each model specification that appears in the union of model confidence sets.

5. Conclusion

In the absence of clear theoretical guidance on specific estimable forms for the aggregate GDP-temperature relationship, we consider the implications of model uncertainty for market damages of climate change. Out-of-sample predictive accuracy is assessed for 800 variants of prominent models that vary by specification of GDP growth or levels effects, as well as by specification of temperature and precipitation functions and controls for unobserved trends. Cross validation is employed to determine model prediction errors, which are used to ascertain the sets of models that have statistically significant superior performance. Across a large subset of models, predictive accuracy is not significantly dependent on the functional form of temperature or its exclusion.

Modeling uncertainty, uniquely estimated in this paper, is shown to rival sampling uncertainty. Growth models generate considerably greater uncertainty of climate impacts than do levels models. For growth models with superior performance, i.e., those contained in any model confidence set (Hansen et al. 2011), the 95% confidence region of GDP impacts in 2100 is −84% to +359%, reflecting considerable model and sampling uncertainty. Accounting for the uncertainty reflected in the set of superior models, we do not identify a statistically significant marginal effect of temperature on global GDP growth.

Though we cannot preclude at 95% confidence that a model relating temperature to GDP growth is superior, we identify models relating temperature to GDP levels as more often being the most accurate in out-of-sample validation. Moreover, growth models incorporating lagged temperature effects indicate that contemporaneous harm from positive temperature shocks are offset in subsequent periods, indicative of temperature effects on GDP levels rather than growth.

Models relating temperature to GDP levels yield climate impact estimates that are far more certain. The best such models imply GDP losses by 2100 of 1–3%, consistent with damage functions currently embedded in the major integrated assessment models that underpin the U.S. social cost of carbon (National Academies of Sciences 2017; Nordhaus 2017; Rose et al. 2017; National Research Council 2010). The 95% confidence range for GDP levels models in any model confidence set is −8.5% to +1.8%. Hot temperatures are estimated to cause statistically significant losses to the *level* of poor country and agricultural GDP, but not to rich-country and non-agricultural production.

While the climate change impacts estimated by GDP-levels models may appear modest, even a 1% loss to global GDP is equal to $800 billion today and could be 5–12 times greater by 2100 amid 2–3% annual economic growth. Moreover, projected GDP impacts based on past temperature fluctuations reflect only a component of potential welfare effects, excluding, for instance, effects on non-market goods like environmental amenities and potential extreme events not reflected in the historical record.

As this analysis has demonstrated the sensitivity of estimated temperature impacts on GDP to both model and sampling uncertainty, it suggests further research is warranted to improve understanding of the relationship between climate and production, particularly at a disaggregated sectoral level. In the spirit of the robust literature on crop impacts of climate change, future work should disaggregate annual temperature data to explore heterogeneous temperature sensitivity across seasons or months of the year, as well as non-linearities in the effects of daily temperatures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jeem.2021.102445.

References

Acemoglu, D., Johnson, S., Robinson, J.A., 2002. Reversal of fortune: geography and institutions in the making of the modern world income distribution. Q. J. Econ. 117 (4), 1231–1294.
Anttila-Hughes, J.K., Hsiang, S.M., 2011. Destruction, Disinvestment, and Death: Economic and Human Losses Following Environmental Disaster.
Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4, 40–79.
Athey, S., 2017. The impact of machine learning on economics. In: Economics of Artificial Intelligence. University of Chicago Press.
Athey, S., Bayati, M., Imbens, G., Qu, Z., 2019. Ensemble methods for causal effects in panel data settings. AEA Papers and Proceedings 109, 65–70.
Auffhammer, M., 2018. Quantifying economic damages from climate change. J. Econ. Perspect. 32 (4), 33–52.
Auffhammer, M., Steinhauser, R., 2012. Forecasting the path of U.S. CO_2 emissions using state-level information. Rev. Econ. Stat. 94 (1), 172–185.
Barrios, S., Bertinelli, L., Strobl, E., 2010. Trends in rainfall and economic growth in Africa: a neglected cause of the African growth tragedy. Rev. Econ. Stat. 92 (2), 350–366.
Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. Rev. Econ. Stud. 81 (2), 608–650.
Burke, M., Davis, W.M., Diffenbaugh, N.S., 2018. Large potential reduction in economic damages under UN mitigation targets. Nature 557 (7706), 549.
Burke, M., Hsiang, S.M., Miguel, E., 2015. Global non-linear effect of temperature on economic production. Nature 527 (7577), 235–239.
Cachon, G.P., Gallino, S., Olivares, M., 2012. Severe Weather and Automobile Assembly Productivity. Columbia Business School Research Paper, (12/37).
Carleton, T.A., Hsiang, S.M., 2016. Social and economic impacts of climate. Science 353 (6304), aad9837.
Chatfield, C., 1996. Model uncertainty and forecast accuracy. J. Forecast. 15 (7), 495–508.
Christensen, Garret, Edward, Miguel, 2018. Transparency, reproducibility, and the credibility of economics research. J. Econ. Lit. 56 (3), 920–980. https://www.aeaweb.org/articles?id=10.1257/jel.20171350
Cooley, T.F., LeRoy, S.F., 1981. Identification and estimation of money demand. Am. Econ. Rev. 71 (5), 825–844.
Dell, M., Jones, B.F., Olken, B.A., 2012. Temperature shocks and economic growth: evidence from the last half century. Am. Econ. J. Macroecon. 4 (3), 66–95.
Dell, M., Jones, B.F., Olken, B.A., 2014. What do we learn from the weather? The new climate–economy literature. J. Econ. Lit. 52 (3), 740–798.
Deryugina, T., 2011. The Dynamic Effects of Hurricanes in the U.S.: the Role of Non-disaster Transfer Payments. (MIT Center for Energy and Environmental Policy Research, Working Paper).
Deryugina, T., Hsiang, S.M., 2014. Does the Environment Still Matter? Daily Temperature and Income in the United States. NBER Working Paper No. 20750.
Deryugina, T., Hsiang, S.M., 2017. The Marginal Product of Climate. NBER Working Paper No. 24072.
Deschênes, O., Greenstone, M., 2007. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. Am. Econ. Rev. 97 (1), 354–385.
Deschênes, O., Greenstone, M., 2012. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: Reply. Am. Econ. Rev. 102 (7), 3761–3773.
Dhrymes, P.J., Howrey, E.P., Hymans, S.H., Kmenta, J., Leamer, E.E., Quandt, R.E., Ramsey, J.B., Shapiro, H.T., Zarnowitz, V., 1972. Criteria for evaluation of econometric models. In: Annals of Economic and Social Measurement, ume 1, pp. 291–324 number 3, NBER.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263.
Easterly, W., Levine, R., 2003. Tropics, germs, and crops: how endowments influence economic development. J. Monetary Econ. 50 (1), 3–39.
Elliott, G., Timmermann, A., 2016. Forecasting in economics and finance. Annual Review of Economics 8, 81–110.
Fankhauser, S., 1994. The social costs of greenhouse gas emissions: an expected value approach. Energy J. 157–184.
Fankhauser, S., 2013. Valuing Climate Change: the Economics of the Greenhouse. Routledge, London.
Feng, S., Krueger, A.B., Oppenheimer, M., 2010. Linkages among climate change, crop yields and Mexico-US cross-border migration. Proc. Natl. Acad. Sci. Unit. States Am. 107 (32), 14257–14262.

Fisher, A.C., Hanemann, W.M., Roberts, M.J., Schlenker, W., 2012. The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather: Comment. Am. Econ. Rev. 102 (7), 3749–3760.
Friedman, M., 1953. The methodology of positive economics. In: Essays in Positive Economics. University of Chicago Press, pp. 3–16.
Geisser, S., 1975. The predictive sample reuse method with applications. J. Am. Stat. Assoc. 70 (350), 320–328.
Graff Zivin, J., Neidell, M., 2014. Temperature and the allocation of time: implications for climate change. J. Labor Econ. 32 (1), 1–26.
Green, W.H., 2012. Econometric Analysis. Prentice Hall, Boston.
Greenstone, M., Kopits, E., Wolverton, A., 2013. Developing a social cost of carbon for us regulatory analysis: a methodology and interpretation. Rev. Environ. Econ. Pol. 7 (1), 23–46.
Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23 (4), 365–380.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Hendry, D.F., 1980. Econometrics-alchemy or Science? *Economica*, pp. 387–406.
Hendry, D.F., Leamer, E.E., Poirier, D.J., 1990. A conversation on econometric methodology. Econom. Theor. 6 (2), 171–261.
Hsiang, S., 2016. Climate econometrics. Annual Review of Resource Economics 8 (1), 43–75.
Hsiang, S.M., 2010. Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. Proc. Natl. Acad. Sci. Unit. States Am. 107 (35), 15367–15372.
Hsiang, S.M., Jina, A.S., 2014. The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth: Evidence from 6,700 Cyclones. NBER Working Paper No. 20352.
Hsiang, S.M., Narita, D., 2012. Adaptation to cyclone risk: evidence from the global cross-section. Climate Change Economics 3 (2), 1250011.
Jones, B.F., Olken, B.A., 2010. Climate shocks and exports. Am. Econ. Rev. 100 (2), 454–459.
Kahn, M.E., 2005. The death toll from natural disasters: the role of income, geography, and institutions. Rev. Econ. Stat. 87 (2), 271–284.
Kaufmann, R.K., Kauppi, H., Stock, J.H., 2010. Does temperature contain a stochastic trend? Evaluating conflicting statistical results. Climatic Change 101 (3–4), 395–405.
Keynes, J.M., 1939. Professor Tinbergen's method. Econ. J. 49 (5), 559–577.
Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Prediction policy problems. Am. Econ. Rev. 105 (5), 491–495.
Koopmans, T.C., 1947. Measurement without theory. Rev. Econ. Stat. 29 (3), 161–172.
Kopits, E., Marten, A., Wolverton, A., 2014. Incorporating 'catastrophic' climate change into policy analysis. Clim. Pol. 14 (5), 637–664.
Leamer, E.E., 1978. Specification Searches: Ad Hoc Inference with Nonexperimental Data, vol. 53. John Wiley & Sons Incorporated.
Leamer, E.E., 1983. Let's take the con out of econometrics. Am. Econ. Rev. 73 (1), 31–43.
Leamer, E.E., 2010. Tantalus on the road to asymptopia. J. Econ. Perspect. 24 (2), 31–46.
Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate trends and global crop production since 1980. Science 333 (6042), 616–620.
Maddison, D., 2003. The amenity value of the climate: the household production function approach. Resour. Energy Econ. 25 (2), 155–175.
Matsuura, K., Willmott, C.J., 2012. Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1900 - 2010) V 3.01. Available at: http://climate.geog.udel.edu/~climate/html_pages/download.html#T2011.
McIntosh, C.T., Schlenker, W., 2006. Identifying Non-linearities in Fixed Effects Models. (UC-San Diego Working Paper).
Mendelsohn, R., Morrison, W., Schlesinger, M.E., Andronova, N.G., 2000a. Country-specific market impacts of climate change. Climatic Change 45 (3–4), 553–569.
Mendelsohn, R., Nordhaus, W.D., Shaw, D., 1994. The Impact of Global Warming on Agriculture: A Ricardian Analysis, pp. 753–771 The American Economic Review.
Mendelsohn, R., Schlesinger, M., Williams, L., 2000b. Comparing impacts across climate models. Integrated Assess. 1 (1), 37–48.
Moore, F.C., Diaz, D.B., 2015. Temperature impacts on economic growth warrant stringent mitigation policy. Nat. Clim. Change 5 (2), 127–131.
Müller, U.K., Stock, J.H., Watson, M.W., 2019. An Econometric Model of International Long-Run Growth Dynamics. Working Paper 26593, National Bureau of Economic Research.
National Academies of Sciences, 2017. Valuing Climate Damages: Updating Estimation of the Social Cost of Carbon Dioxide. National Academies Press.
National Research Council, 2010. Hidden Costs of Energy: Unpriced Consequences of Energy Production and Use. National Academies Press.
Nordhaus, W.D., 1994. Managing the Global Commons: the Economics of Climate Change, vol. 31. MIT Press, Cambridge, MA.
Nordhaus, W.D., 2006. Geography and macroeconomics: new data and new findings. Proc. Natl. Acad. Sci. U.S.A. 103 (10), 3510–3517.
Nordhaus, W.D., 2008. A Question of Balance: Weighing the Options on Global Warming Policies. Yale University Press, New Haven.
Nordhaus, W.D., 2017. Revisiting the social cost of carbon. Proc. Natl. Acad. Sci. Unit. States Am. 114 (7), 1518–1523.
Nordhaus, W.D., Moffat, A., 2017. A Survey of Global Impacts of Climate Change: Replication, Survey Methods, and a Statistical Analysis. NBER Working Paper No. 23646.
O'Neill, B.C., Kriegler, E., Riahi, K., Ebi, K.L., Hallegatte, S., Carter, T.R., Mathur, R., van Vuuren, D.P., 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Climatic Change 122 (3), 387–400.
Opsomer, J., Wang, Y., Yang, Y., 2001. Nonparametric regression with correlated errors. Stat. Sci. 16 (2), 134–153.
Ortiz-Bobea, A., Knippenberg, E., Chambers, R.G., 2018. Growing climatic sensitivity of us agriculture linked to technological change and regional specialization. Science Advances 4 (12), eaat4343.
Poterba, J.M., 1993. Global warming policy: a public finance perspective. J. Econ. Perspect. 7 (4), 47–63.
Raftery, A.E., Zimmer, A., Frierson, D.M.W., Startz, R., Liu, P., 2017. Less than 2°c warming by 2100 unlikely. Nat. Clim. Change 7 (9), 637–641.
Ricke, K., Drouet, L., Caldeira, K., Tavoni, M., 2018. Country-level social cost of carbon. Nat. Clim. Change 8 (10), 895–900.
Rodrik, D., Subramanian, A., Trebbi, F., 2004. Institutions rule: the primacy of institutions over geography and integration in economic development. J. Econ. Growth 9 (2), 131–165.
Rose, S.K., Diaz, D.B., Blanford, G.J., 2017. Understanding the social cost of carbon: a model diagnostic and inter-comparison study. Climate Change Economics 8 (2), 1750009.
Schlenker, W., Auffhammer, M., 2018. The cost of a warming climate. Nature 557 (3–4), 498–499.
Schlenker, W., Lobell, D.B., 2010. Robust negative impacts of climate change on African agriculture. Environ. Res. Lett. 5 (1), 014010.
Schlenker, W., Roberts, M.J., 2006. Nonlinear effects of weather on corn yields. Appl. Econ. Perspect. Pol. 28 (3), 391–398.
Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. In: Proceedings of the National Academy of Sciences, vol. 106, pp. 15594–15598 37.
Snee, R.D., 1977. Validation of regression models: methods and examples. Technometrics 19 (4), 415–428.
Stern, N., 2006. The Economics of Climate Change: the Stern Review. Cambridge University Press.
Stevens, A., 2017. Temperature, Wages, and Agricultural Labor Productivity. (Working Paper).
Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P., 2013. IPCC, 2013: Summary for Policymakers in Climate Change 2013: the Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.
Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. Roy. Stat. Soc. B 111–147.
Sudarshan, A., Somanathan, E., Somanathan, R., Tewari, M., 2015. The Impact of Temperature on Productivity and Labor Supply - Evidence from Indian Manufacturing. Centre for Development Economics, Dehli School of Economics Working Paper No. 244.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. J. Finance 54 (5), 1647–1691.
Tol, R.S., 1995. The damage costs of climate change toward more comprehensive calculations. Environ. Resour. Econ. 5 (4), 353–374.
Tol, R.S., 2002a. Estimates of the damage costs of climate change. Part 1: benchmark estimates. Environ. Resour. Econ. 21 (1), 47–73.
Tol, R.S., 2002b. Estimates of the damage costs of climate change, Part II. Dynamic estimates. Environ. Resour. Econ. 21 (2), 135–160.
Tol, R.S., 2014. Correction and update: the economic effects of climate change. J. Econ. Perspect. 28 (2), 221–226.
Tol, R.S.J., 2009. The economic effects of climate change. J. Econ. Perspect. 23 (2), 29–51.
van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.J., Rose, S.K., 2011. The representative concentration pathways: an overview. Climatic Change 109 (1–2), 5.
Varian, H.R., 2014. Big data: new tricks for econometrics. J. Econ. Perspect. 28 (2), 3–28.
West, K.D., 1996. Asymptotic inference about predictive ability. Econometrica: Journal of the Econometric Society 64 (5), 1067–1084.
White, H., 1996. Estimation, Inference and Specification Analysis, vol. 22. Cambridge University Press.
World Bank, 2012. World Development Indicators 2012. World Bank, Washington, U.S..
Yatchew, A., 1998. Nonparametric regression techniques in economics. J. Econ. Lit. 36 (2), 669–721.