Reckoning Climate Change Damages along an Envelope

Matthew Gammans[⋆], Pierre Mérel[†] and Emmanuel Paroissien[‡]

[⋆]Agricultural, Food, and Resource Economics, Michigan State University, 446 West Circle Drive, East Lansing, MI 48824

[†]Agricultural and Resource Economics, University of California, Davis, 1 Shields Avenue, Davis CA 95616

[‡]Université Paris-Saclay, INRAE, UR ALISS, 94205 Ivry-sur-Seine, France

Abstract

We derive necessary and sufficient conditions for weather fluctuations to systematically identify the marginal long-run effect of climate on an optimized outcome. Empirical estimates of local marginal effects trace out a common long-run response function that can be used for non-marginal climate counterfactual analysis. Our empirical application considers the effect of temperature and precipitation on county-level agricultural GDP in the United States. Except in our most demanding specification, which delivers imprecise estimates, agricultural GDP is predicted to decrease significantly due to warming.

JEL codes: C23, Q16, Q54

Keywords: climate change, climate adaptation, envelope theorem, weather

1 Introduction

Since the seminal work of Mendelsohn et al. (1994), economists have sought ways to predict the impacts of climate change on economic outcomes using observational data. Damage estimates derived from econometric models can be informative by themselves, but they are also needed to calibrate larger models of the economy (Auffhammer, 2018). One of the main empirical challenges in exploiting historical patterns of climate or weather and economic outcomes is attribution. In an ideal but unrealistic situation, one could observe two otherwise identical economies subject to different climates, and track relevant outcomes.

By necessity, empirical work has departed from this ideal comparison in two main directions. Keeping with the core idea that the relevant counterfactual comparison involves different climates, Mendelsohn et al. (1994) and the ensuing "Ricardian" literature have compared outcomes across locations exposed to different climates, controlling to the extent possible for potential cross-sectional confounders (Schlenker et al., 2005; Ortiz-Bobea, 2020). The belief that no amount of care would dissipate concerns about omitted variable bias has led another branch of literature to favor panel approaches with fixed effects, which exploit assumedly random weather fluctuations as a source of identifying variation (Deschênes and Greenstone, 2007).[1] The main weakness of that approach is that the effect of weather is conceptually different from the effect of climate due to plausible adaptation (Hsiang, 2016; Auffhammer, 2018).[2]

The present paper proposes a novel theoretical and empirical framework that can arguably make the best out of both worlds. Our main empirical objects are a collection of local marginal effects identified purely from random weather fluctuations using historical data. We use these objects to trace out a common long-run climatic response function without any parametric restriction beyond that inherent in the choice of relevant climatic variables, making use of the Envelope Theorem and the Gradient Theorem, as initially suggested by Hsiang (2016). The fact that we rely on these theorems for identification of climatic effects does put restrictions on the applicability of our framework, however.

We begin by characterizing formally the instances whereby the Envelope Theorem may be legitimately invoked to claim that the long-run, adaptation-inclusive marginal effects of permanent changes in environmental conditions (say, climate as measured over several decades) on an outcome of interest are identical to the marginal effects of transient, adaptation-exclusive changes in these conditions (say, yearly weather). In doing so, we introduce a new nomenclature meant to clarify the discourse about the use of the Envelope Theorem in climate impact assessment by defining and relating

[1]A third approach, explored in Burke and Emerick (2016), exploits county-level climatic trends, residual of state-level trends, between distant periods.

[2]Nonlinear panel approaches exploit both weather variation and climatic differences, and can be expected to deliver a response that is a mix of short-run and long-run effects (Kolstad and Moore, 2020). Within a quadratic framework, Mérel and Gammans (2021) formally show that the identified relationship is a convex combination of these effects, and that the approach is not immune to omitted variable bias if unobserved climate correlates interact with weather.

various outcome response functions, notably that inferable from observational data and those that may be used to build relevant climate counterfactuals. We show that in addition to the usual smoothness assumption and the fact that the outcome of interest must be an optimized value, exploiting the envelope property requires certain functional restrictions on the relationship between weather, climatic adaptation, and the outcome of interest. These restrictions relate directly to Theil (1954)'s sufficient conditions regarding the use of expectations in social welfare maximization problems.[3] A substantial contribution of our paper is to prove that these conditions are not only sufficient, as argued by Theil (1954), but also necessary.

We then demonstrate how the envelope property can be leveraged empirically to trace out a global long-run response to climate without imposing any parametric restrictions on the shape of the response function, while addressing concerns about omitted variables. Intuitively, we exploit assumedly random fluctuations in weather at each given climate value to estimate a local slope, which reflects both short-run and long-run effects under the tangency property implied by the Envelope Theorem. Riemann integration of local slopes estimates across adjacent climates then delivers the global long-run response function.

Our empirical implementation of this envelope-gradient-theorem approach uses agricultural GDP observed over 17 years across more than a thousand US counties. In spite of the high flexibility of our approach and the demands put on our data to estimate climate-specific marginal responses, the long-run responses to climatic variables that we obtain are surprisingly smooth and precisely estimated. In addition, they are coherent with stylized expectations as well as basic restrictions imposed in prior parametric work on temperature and agricultural outcomes such as yields or farmland values.

There is a perception in the literature that the principal restrictions to the use of the Envelope Theorem pertain to whether the outcome is optimized and whether adaptations can be considered to have a continuous, rather than discrete, effect on the outcome (Guo and Costello, 2013; Nordhaus, 2010; Hsiang, 2016; Blanc and Schlenker, 2017; Deryugina and Hsiang, 2017).[4] Here, we argue that even for optimized outcomes that depend smoothly on adaptation actions, the Envelope Theorem does not necessarily imply that random weather variation can be used to identify the slope of

[3]Theil (1954) is concerned with the use of expectations of random variables in static, nonlinear welfare maximization programs in place of full distributions of these variables. Simon (1956) and Theil (1957) address a similar issue in the context of dynamic programming.

[4]In related work, Lemoine (2021) shows how the envelope-theorem argument can also fail in a dynamic framework.

the long-run outcome function with respect to climate. The reason is that tangency between long-run and short-run responses at the mean weather (i.e., the first moment of climate) is guaranteed only when long-run actions are taken to optimize outcomes under such mean weather. More realistically however, long-run actions are taken to optimize *expected* outcomes under the entire weather *distribution*, as recognized in several studies (Kelly et al., 2005; Schlenker, 2017; Shrader, 2021; Maue and Kolstad, 2020; Lemoine, 2021; Mérel and Gammans, 2021). As a result, even under risk neutrality the slopes (and the values) of the short-run and long-run responses may not coincide at the mean weather. If not, they coincide at some other weather value, but without further functional assumptions it is impossible to know where, which renders the tangency property unexploitable in practice.[5] Thus, without a proper tangency condition, random weather fluctuations are of little use if analysts care about long-run responses that include adaptation, which is almost always the case (Auffhammer, 2018).

Having made that point formally, our paper goes on to derive structural restrictions on the outcome function that ensure that even under expected outcome maximization, an envelope result systematically holds at mean weather, allowing the identification of long-run effects of environmental change from random high-frequency fluctuations. These conditions are shown to be both necessary and sufficient within a broad class of smooth and convex optimization problems assumed to underlie long-run adaptation choices. Like previous authors, we assume that observed outcomes $y(x, a)$ depend structurally on a realization of weather x and an unobserved action (or set of actions) a taken in response to climate, and that conditional on the weather realization, climate influences outcomes only through these actions. This structural relationship gives rise to a reduced-form relationship between weather, climate, and the outcome. The key structural restriction implied by the tangency property is that the actions a are allowed to interact only with a *linear* function of weather, that is, $y(x, a) = \Gamma(x) + \Psi(a)x + \Phi(a)$, for some smooth functions Γ, Ψ, and Φ. Said differently, tangency between the short-run response to weather and the long-run response to climate (at mean weather) requires the marginal effect of farmers' actions to depend on weather in a linear fashion, even if the function Γ itself is unrestricted.[6] We then go on to show that the marginal response to weather also identifies the marginal *expected* response to climate if and only if the function Γ is quadratic (or affine). These conditions may

[5]For a parametric framework that explores this possibility, see Schlenker (2017). His approach allows tangency to happen elsewhere than at mean weather, and in a flexible way, but the global relationship is restricted parametrically.

[6]Of course, the functions Ψ and Φ are also unrestricted as they involve an unobserved action, which could be redefined through a change of variable.

seem restrictive, but the key insight of our analysis is that they are needed if one is to exploit the tangency property implied by the Envelope Theorem to recover marginal climate effects from the observation of weather effects. In fact, a comparable tradeoff is described by Theil (1954) in the context of the social welfare maximization problem under uncertainty.[7]

Armed with a deeper understanding of what the envelope property, in its most usable form, implicitly assumes, we proceed with its empirical implications. The main consequence of local tangency is that, for the purpose of delineating marginal climate effects, weather fluctuations can be considered as good as climatic changes. Because the assumption of exogeneity is much more likely to hold for weather than for climate, this property opens the door to consistent estimation of marginal climate effects, which can then be integrated across adjacent climates to obtain a common long-run response function. In our application, we regress annual measures of US county-level agricultural GDP on average temperature and cumulative precipitation. For each climate variable, we estimate marginal effects for 100 climatic intervals. These slopes are then used to construct the long-run response functions of agricultural GDP to temperature and precipitation, free of functional form restrictions.

Our framework also entails a potential improvement over previous empirical work in terms of model identification. Different regions may exhibit distinct responses to weather based on geographic variables, such as soil type or elevation, or based on institutional and macroeconomic disparities. To the extent that these variables are correlated with climate, the climate-specific marginal effects we identify may be biased as they would reflect the influence of these factors in addition to that of climate. Because climates have changed during the period of observation, in our most conservative specification we choose to control for some of these time-invariant factors through the inclusion of state-specific weather slopes in the estimation of marginal weather responses at each climate interval. Consequently, any contribution of a state's time-invariant unobservables to its marginal weather response is removed from identi-

[7]Specifically, Theil (1954) writes the following:

> The main conclusion is that there is a central class of cases for which the short-cut is permissible. If [...] then the short-cut is bound to lead to the same policy decisions as those taken in the case of full information. The important thing is that this result does not require any additional specifications of the policy maker's preferences. In other cases the short-cut will usually lead to choice bias. It is in principle possible to avoid this by replacing the stochastic vector [...] by a non stochastic vector different from the expectation, but such a short-cut requires specifications of the welfare function that go much farther than [our] assumption.

fication, which then relies in part on counties crossing climatic intervals over time due to secular warming or cooling. That is, differences in weather slopes across climatic intervals are identified from comparing a county to itself under different climates, and from comparing counties with different climates but within the same state.

The rest of the paper proceeds as follows. Section 2 investigates the theoretical underpinnings of the envelope property. Section 3 discusses its consequences in terms of empirical identification of climate effects. Section 4 presents our empirical application, and Section 5 concludes.

2 The envelope argument: what are we assuming?

Previous literature has already acknowledged two central assumptions underlying the use of the Envelope Theorem, namely differentiability (Guo and Costello, 2013) and the fact that outcomes must be optimized (Hsiang, 2016). Here, we argue that in addition to these caveats, functional restrictions are in fact needed to usefully invoke the envelope property as a basis for the empirical identification of a marginal long-run response to climate from random weather fluctuations. To our knowledge, we are the first to make these restrictions explicit.

For clarity's sake, we consider unidimensional weather, denoted by $x \in \mathbb{R}$. We relax this assumption in our empirical application. We denote by F the cumulative distribution function of x and call it the *climate*. The expectation of weather is denoted by $\mathbb{E}[x] = \mu$. In many empirical studies it is the only aspect of climate taken into consideration.[8]

The outcome is a structural function $y(x, a)$ of weather and an action $a \in U \subset \mathbb{R}$ taken prior to (or without knowledge of) the realization of weather, where U denotes an open, convex subset of $\mathbb{R}$. Importantly, conditional on the weather realization x, the value of the outcome only depends on climate (or its moments) through the action a. If multiple actions can be taken, then a is interpretable as an index (mapping) that varies according to this set of actions. Hence, the assumption of a univariate action does not unduly restrict the analysis. We call $y(x, a)$ the *structural outcome function*.

2.1 Regularity assumptions

We restrict the function y to be continuously differentiable with respect to x and a, to be concave with respect to a, and to have a unique maximizer in the a dimension, for

[8] Schlenker (2017) is a notable exception.

given x. As a result, for all values of x, this unique maximizer, to be denoted $\hat{a}(x)$, is characterized by the first-order condition

$$\frac{\partial y}{\partial a}(x, a) = 0.$$

The solution to the previous equation, $\hat{a}(x)$, thus answers the question: "what action would be optimal to maximize the outcome under weather x?" Note that the concavity assumption is made here without loss of generality once we require that there be a unique critical point characterizing the global maximum, as one can always redefine the index a through a well-chosen change of variable such that the resulting function be concave. (For a formal proof of this claim, see Appendix A.)

We further assume that the function $\hat{a}(x)$ is surjective, though not necessarily bijective, on U, that is, each possible action $a \in U$ is optimal for some value(s) of weather. This implies that the equation $\frac{\partial y}{\partial a}(x, a) = 0$ has at least one solution $x \in \mathbb{R}$ for any given value $a \in U$. Solving this equation for x rather than a answers the question: "for what value(s) of weather would the action a be optimal?"

We also make the following assumption:

Assumption 1 *The function $\frac{\partial y}{\partial x}(x, a)$ is injective with respect to a, that is, $\frac{\partial y}{\partial x}(x, a) = \frac{\partial y}{\partial x}(x, a') \Rightarrow a = a'$.*

Assumption 1 requires that the marginal effect of weather on the outcome depend on the action taken, in the sense that different actions imply different marginal effects for given weather. Note that this assumption does not preclude the same action a from being optimal for different values of weather, say x and x', as the corresponding conditions would be $\frac{\partial y}{\partial a}(x, a) = \frac{\partial y}{\partial a}(x', a) = 0$.

Finally, note that the fact that the action a is taken without knowledge of the weather realization does not mean that there is no possibility of adaptation to weather once realized; if there is, the effects of such adaptation are simply accounted for in the structural outcome function $y(x, a)$. That is, $y(x, a)$ can be thought of as the value function of an optimization problem where adaptation to weather is assumed to maximize the outcome, conditional on weather x and the long-run action a.

2.2 Statement of the problem

We define the *long-run outcome response function* to climate μ as the value of the following problem:[9]

$$\left\{\max_{a\in U}\, y(\mu,a)\right\} = y\left(\mu,\hat{a}(\mu)\right) \equiv Y(\mu). \tag{1}$$

This definition conforms with the concept of "long run" as typically used in microeconomics textbooks, e.g., Perloff (2016) or Nicholson and Snyder (2016). For instance, although firms cannot vary capital in each production period, the long-run cost function is one that minimizes the cost of producing any given quantity when allowing all factors, including capital, to vary with output. This long-run cost function tells us how production costs would change if output quantity were to change from say, durably low levels to durably high levels. However, it is not very informative if we are interested in the effects of transient production shocks, because it assumes a degree of flexibility that the firm typically does not have. Similarly, the function $Y(\cdot)$ indicates how the outcome *would* respond to weather, if the action a *could* be taken in anticipation of weather. It is informative about how the outcome would respond if climate changed from say, a durably cool climate to a durably warm climate.

The Envelope Theorem implies that $Y'(\mu) = \frac{\partial y}{\partial x}\left(\mu,\hat{a}(\mu)\right)$. This theorem has been invoked in the literature as a way to justify the use of weather fluctuations to identify the marginal effects of changes in climate. The reason is that it is easier to observe weather shocks than climatic shocks in empirical work. Weather changes at a yearly frequency, sometimes with large swings, providing arguably exogenous variation in many settings. In contrast, climate may be stationary, or may only change at a slow pace over time. These changes may only be detectable over very long periods of time during which other relevant factors may change, raising concerns about omitted variable bias.

Unfortunately, the function observed in the data may not be $y\left(x,\hat{a}(\mu)\right)$, as implicitly assumed by the use of the Envelope Theorem. For $\hat{a}(\mu)$ is the long-run adaptation that *maximizes the outcome under mean weather*. Instead, if economic agents are taking long-run actions in response to climatic signals, a more tenable assumption is that they choose a to maximize *expected outcomes*, that is, the expected realization of y given the climate F. The same point has been made, for instance, in contributions by Kelly et al. (2005), Schlenker (2017), Shrader (2021), or Carleton et al. (2020). Conceptually, agents'

[9]The squiggly brackets around the maximization problem, although perhaps redundant, are meant to emphasize that the element on the right-hand side of the equality sign is equal to the value, rather than the objective, function. The equivalency sign (≡) indicates a definition.

Table 1 Definitions of functions

Function	Name
$y(x,a)$	*structural outcome function*
$\hat{y}(x,\mu) \equiv y\left(x,\hat{a}(\mu)\right)$	*conditional reduced-form outcome function*
$Y(\mu) \equiv \hat{y}(\mu,\mu)$	*long-run outcome response function*
$\tilde{y}(x,F) \equiv y\left(x,\tilde{a}(F)\right)$	*short-run reduced-form outcome function*
$\mathcal{Y}(F) \equiv \int \tilde{y}\left(x,F\right)dF(x)$	*long-run expected outcome response function*

choice of action, denoted $\tilde{a}(F)$, would thus solve the following problem:

$$\left\{\max_{a\in U}\int y(x,a)dF(x)\right\} = \int y\left(x,\tilde{a}(F)\right)dF(x). \tag{2}$$

There is a conceptual difference between maximizing the outcome under mean weather, as in Problem (1), and maximizing the expectation of the outcome under given climate, as in Problem (2). Notably, maximization of expected outcomes implies that higher-order moments of the climate, for instance its variance, may influence choices, whereas only the mean matters in Problem (1).

We will denote $\hat{y}(x,\mu) \equiv y\left(x,\hat{a}(\mu)\right)$ the reduced-form outcome function that would obtain if a were chosen to solve Problem (1). The long-run outcome response function is related to this reduced-form outcome function because $Y(\mu) = y\left(\mu,\hat{a}(\mu)\right) = \hat{y}(\mu,\mu)$, and $Y'(\mu) = \frac{\partial\hat{y}}{\partial x}(\mu,\mu)$ since $\frac{\partial\hat{y}}{\partial\mu}(\mu,\mu) = \frac{\partial y}{\partial a}\left(\mu,\hat{a}(\mu)\right)\hat{a}'(\mu) = 0$ from the first-order condition to Problem (1). In contrast, we will denote $\tilde{y}(x,F) \equiv y\left(x,\tilde{a}(F)\right)$ the reduced-form outcome function that obtains when a is chosen to solve Problem (2). It is this latter function, not $\hat{y}(x,\mu)$, that we expect to observe in the data. We will thus refer to $\tilde{y}(x,F)$ as the *short-run reduced-form outcome function*, and to $\hat{y}(x,\mu)$ as the *conditional reduced-form outcome function*.

Unfortunately, the relationship between the functions $\tilde{y}(x,F)$ and $Y(\mu)$ is not as straightforward as that between $\hat{y}(x,\mu)$ and $Y(\mu)$. Due to Assumption 1, the slopes $\frac{\partial y}{\partial x}\left(\mu,\hat{a}(\mu)\right)$ and $\frac{\partial y}{\partial x}\left(\mu,\tilde{a}(F)\right)$ differ unless $\hat{a}(\mu) = \tilde{a}(F)$. If $\hat{a}(\mu) \neq \tilde{a}(F)$ and agents solve Problem (2) rather than (1), then the Envelope Theorem, which is based on Problem (1), cannot be legitimately invoked in empirical work. That is, a local slope obtained using random weather fluctuations (i.e., $\frac{\partial\tilde{y}}{\partial x}\left(\mu,F\right)$) does not identify the underlying long-run slope $Y'(\mu) = \frac{\partial\hat{y}}{\partial x}(\mu,\mu)$.

In addition to the long-run outcome response function $Y(\mu)$, one may be interested in what we may call the *long-run expected outcome response function*, defined as

$$\mathcal{Y}(F) \equiv \int y\left(x, \tilde{a}(F)\right) dF(x) = \int \tilde{y}\left(x, F\right) dF(x).$$

This function recognizes that economic agents decide on actions a based on expectations, but that once actions are set optimally outcomes are still random due to the randomness of weather. Said differently, $\mathcal{Y}(F)$ is the expectation of the short-run reduced-form relationship $\tilde{y}(x, F)$ that we expect to observe in the data, and it is the value function of Program (2). Can the Envelope Theorem be used here to argue that the marginal effect of climate on the expected outcome, $\frac{\partial \mathcal{Y}}{\partial \mu}$, can be identified from the empirically identified marginal change in weather, namely $\frac{\partial \tilde{y}}{\partial x}\left(\mu, F\right)$? Unfortunately the answer is still no, because taking the partial derivative of the objective function in Program (2) with respect to the mean weather μ (holding actions a constant at $\tilde{a}(F)$) would entail taking the derivative of the probability density function, rather than that of the function $y(x, a)$ itself. Intuitively, the objective function of Program (2) is an average, while the function we observe in the data is a particular realization.

Table 1 summarizes the definitions of the various functions discussed above. In what follows, we first derive a necessary and sufficient condition on the structural outcome function $y(x, a)$ under which $\tilde{a}(F) = \hat{a}(\mu)$ for all F, where μ denotes the expected weather under climate F.[10] This is, in essence, the same problem as that analyzed by Theil (1954) in the context of the social welfare maximization problem under uncertainty, although his specification of the structural outcome function is slightly different from ours and he does not attempt to prove necessity. While Theil (1954)'s interest lies chiefly in identifying restrictions under which a policymaker may use expected values in lieu of distribution functions in the expected welfare maximization program, in our context the property $\tilde{a}(F) = \hat{a}(\mu)$ further implies that the marginal effect of weather, evaluated at the mean weather, identifies the long-run marginal effect of climate, that is, $\frac{\partial \tilde{y}}{\partial x} = Y'$. In addition to this key property, we derive a more restrictive necessary and sufficient condition under which the marginal effect of weather also identifies the marginal effect of climate on the long-run expected outcome, that is, $\frac{\partial \tilde{y}}{\partial x} = \frac{\partial \mathcal{Y}}{\partial \mu}$.

[10] Alternatively, one could imagine placing restrictions on the set of distributions F for which the tangency property holds. This approach seems less useful to us and lies beyond the scope of this paper.

2.3 Necessary condition for $\tilde{a}(F) = \hat{a}(\mu)$

Suppose that, despite the fact that Problem (1) and Problem (2) are conceptually different, it is nonetheless the case that $\tilde{a}(F) = \hat{a}(\mu)$ for all F, where μ denotes the expectation of F. We now show the following key proposition:

Proposition 1 *The function $\frac{\partial y}{\partial a}(x,a)$ is <u>affine in x</u> for all $a \in U$.*

<u>Proof</u>: Consider some $a \in U$. By assumption, there exists $\mu \in \mathbb{R}$ such that $\frac{\partial y}{\partial a}(\mu, a) = 0$, that is, $a = \hat{a}(\mu)$. Then for all distributions F with mean μ, it must be the case that $\int \frac{\partial y}{\partial a}(x,a)dF(x) = 0$ from the first-order condition of Problem (2). Define the function $\phi_a(x) = \frac{\partial y}{\partial a}(x+\mu, a)$. Then $\phi_a(0) = 0$. It is also the case that for all distributions G with mean zero, $\int \phi_a(x)dG(x) = 0$. To see why, rewrite

$$\int \phi_a(x)dG(x) = \int \frac{\partial y}{\partial a}(x+\mu,a)g(x)dx = \int \frac{\partial y}{\partial a}(x,a)g(x-\mu)dx = 0$$

where g is the p.d.f. associated with G and the last equality obtains because the density $f(x) \equiv g(x-\mu)$ has mean μ. (To see why, note that $\int xf(x)dx = \int xg(x-\mu)dx = \int (x+\mu)g(x)dx = \mu$ since G has mean zero.)

Using the fact that $\int \phi_a(x)dG(x) = 0$ for all G with mean zero, we will now show that ϕ_a is a linear function. First, for all $x \in \mathbb{R}$, we must have that $\frac{\phi_a(x)}{2} + \frac{\phi_a(-x)}{2} = 0$ since $\frac{x}{2} + \frac{-x}{2} = 0$. That is, ϕ_a is an odd function. Second, for all $x > 0$, consider the distribution $G_x = \left(-1, x; \frac{x}{x+1}, \frac{1}{x+1}\right)$ that takes on the value -1 with probability $\frac{x}{x+1}$ and the value x with probability $\frac{1}{x+1}$. G_x is a zero-mean distribution, therefore we must have

$$\frac{x}{x+1}\phi_a(-1) + \frac{1}{x+1}\phi_a(x) = 0$$

that is, $\phi_a(x) = -x\phi_a(-1) = x\phi_a(1)$ where we have used the fact that ϕ_a is odd. Finally, for $x < 0$, we have $\phi_a(x) = -\phi_a(-x) = -(-x)\phi_a(1) = x\phi_a(1)$. Therefore, for all $x \in \mathbb{R}$, we have that $\phi_a(x) = \phi_a(1)x$, that is, ϕ_a is a linear function.

Since $\phi_a(x) = \frac{\partial y}{\partial a}(x+\mu, a)$, we must have $\frac{\partial y}{\partial a}(x,a) = \phi_a(x-\mu) = \phi_a(1)(x-\mu) = \phi_a(1)x - \phi_a(1)\mu$, which shows that the function $\frac{\partial y}{\partial a}$ is affine in x. Q.E.D.

2.4 Implied structural outcome function

Given that $\frac{\partial y}{\partial a}$ is affine in x for all $a \in U$, it must be that

$$y(x,a) = \Gamma(x) + \Psi(a)x + \Phi(a) \tag{3}$$

for some C^1 functions Γ, Ψ, and Φ. Note that the structure in Equation (3) only imposes that actions interact with a *linear* function of weather, but the functions Γ, Ψ, and Φ may themselves be nonlinear. Therefore, if we assume that expected-outcome-maximizing actions are the same as outcome-maximizing actions under mean weather, we essentially impose that actions interact with a linear function of weather in the structural outcome function. Another way to express this restriction would be to say that the marginal effect of weather on the marginal effect of actions is constant with respect to weather.

It is easy to check that for the class of functions $y(x,a)$ defined by Equation (3), $\int \frac{\partial y}{\partial a}(x,a)dF(x) = \frac{\partial y}{\partial a}(\mu,a)$, where μ is the mean of the distribution F. Therefore, if a maximizes $\int y(x,a)dF(x)$, it also maximizes $y(\mu,a)$, hence $\hat{a}(\mu) = \tilde{a}(F)$ and the structure in (3) is not just necessary, but also sufficient.

Therefore, a necessary and sufficient condition for $\frac{\partial \tilde{y}}{\partial x} = Y'$ is that the structural outcome function follow the structure in Equation (3). This result directly relates to the so-called "certainty equivalence" principle in dynamic programming, which finds its origin in the static work of Theil (1954) and was developed by Simon (1956) and Theil (1957). This principle states that under certain functional restrictions, the optimal policy strategy under uncertainty is equivalent to that which would be optimal if all uncertain variables were replaced by their expected value. In Appendix B, we show that the structure in Equation (3) is equivalent to the functional restrictions identified by Theil (1954).

2.5 Long-run expected outcome response function

Here we are interested in the behavior of the function $\mathcal{Y}(F)$, particularly how it changes when the mean of climate changes, that is, when there is a shift in the weather distribution, say towards higher temperatures, without a change in the shape of that distribution. We derive a necessary and sufficient condition under which $\frac{\partial \mathcal{Y}}{\partial \mu} = \frac{\partial \tilde{y}}{\partial x}$, which provides the needed envelope result (equality of slopes).

For expositional convenience, we consider distribution functions F that can be parsimoniously described by two parameters μ and σ^2 reflecting the mean and variance of weather, respectively. Our derivations would still hold if we allowed higher moments to describe the set of possible distributions. We denote $f(x,\mu,\sigma^2)$ the p.d.f. of F. We can then view $\mathcal{Y}(F)$ as a function of the two distributional parameters, say $\mathcal{Y}(\mu,\sigma^2)$. We are interested in the change in the expected outcome with respect to μ, holding σ^2 constant, that is, $\frac{\partial \mathcal{Y}}{\partial \mu}$, and therefore it may be useful to derive conditions under which the empirically identified effect $\frac{\partial \tilde{y}}{\partial x}(\mu, F)$ represents the effect $\frac{\partial \mathcal{Y}}{\partial \mu}$, in addition to representing the effect $Y'(\mu)$ discussed previously. We thus consider the set of structural outcome functions $y(x, a)$ satisfying the structure in Equation (3), and ask what additional restrictions, if any, are needed for the empirically identified marginal effect to also reflect the local behavior of the function $\mathcal{Y}(\mu, \sigma^2)$ with respect to μ.

Given Equation (3), we have that $\tilde{y}(x, \mu, \sigma^2) = \Gamma(x) + \Psi\left(\tilde{a}(\mu, \sigma^2)\right) x + \Phi\left(\tilde{a}(\mu, \sigma^2)\right)$ and therefore $\frac{\partial \tilde{y}}{\partial x}(\mu, \mu, \sigma^2) = \Gamma'(\mu) + \Psi\left(\tilde{a}(\mu, \sigma^2)\right)$. We also have that $\mathcal{Y}(\mu, \sigma^2) = \int \Gamma(x) dF(x) + \Psi(\tilde{a}(\mu, \sigma^2))\mu + \Phi(\tilde{a}(\mu, \sigma^2))$ and therefore

$$\begin{aligned}
\frac{\partial \mathcal{Y}}{\partial \mu}(\mu, \sigma^2) &= \int \Gamma(x) \frac{\partial f}{\partial \mu}(x, \mu, \sigma^2) dx + \Psi(\tilde{a}(\mu, \sigma^2)) \\
&\quad + \left[\Psi'(\tilde{a}(\mu, \sigma^2))\mu + \Phi'(\tilde{a}(\mu, \sigma^2))\right] \frac{\partial \tilde{a}}{\partial \mu}(\mu, \sigma^2) \\
&= \int \Gamma(x) \frac{\partial f}{\partial \mu}(x, \mu, \sigma^2) dx + \Psi(\tilde{a}(\mu, \sigma^2))
\end{aligned}$$

where we have used the fact that $\tilde{a}(\mu, \sigma^2) = \hat{a}(\mu)$ together with the first-order condition for the maximization of $y(\mu, a)$.

Therefore, it is apparent that the essential condition for $\frac{\partial \mathcal{Y}}{\partial \mu} = \frac{\partial \tilde{y}}{\partial x}$, under the structural model implied by (3), is that, for all distributions F, $\Gamma'(\mu) = \int \Gamma(x) \frac{\partial f}{\partial \mu}(x, \mu, \sigma^2) dx$, that is:

$$\Gamma'(\mu) = \frac{\partial}{\partial \mu}\left(\int \Gamma(x) dF(x)\right). \tag{4}$$

Condition (4) is trivially satisfied if $\Gamma(x)$ is affine. It is also satisfied if $\Gamma(x)$ is quadratic. To see why, write $\Gamma(x) = ax^2 + bx + c$. We then have:

$$\begin{aligned}
\int \Gamma(x) dF(x) &= a \int x^2 f(x, \mu, \sigma^2) dx + b \int x f(x, \mu, \sigma^2) dx + c \int f(x, \mu, \sigma^2) dx \\
&= a(\mu^2 + \sigma^2) + b\mu + c
\end{aligned}$$

and therefore $\frac{\partial}{\partial \mu}\left(\int \Gamma(x) dF(x)\right) = 2a\mu + b = \Gamma'(\mu)$.

However, the property does not carry forward if $\Gamma(x)$ is cubic.[11] As expressed in

[11] Consider the function $\Gamma(x) = a'x^3 + b'x^2 + c'x + d'$. Writing $x^3 = (x - \mu + \mu)^3 = (x - \mu)^3 + 3\mu(x - \mu)^2 + 3\mu^2(x - \mu) + \mu^3$ and defining $M_3 \equiv \int (x - \mu)^3 dF(x)$, we obtain:

$$\int \Gamma(x) f(x, \mu, \sigma^2) dx = a'(M_3 + \mu^3 + 3\mu\sigma^2) + b'(\mu^2 + \sigma^2) + c'\mu + d'$$

and therefore $\int \Gamma(x) \frac{\partial f}{\partial \mu}(x, \mu, \sigma^2, M_3) dx = 3a'(\mu^2 + \sigma^2) + 2b'\mu + c'$ which is different from $\Gamma'(\mu) =$

the following proposition, it turns out that the fact that $\Gamma(x)$ be quadratic is necessary for condition (4) to hold for all F.

Proposition 2 *A necessary and sufficient condition for $\frac{\partial \mathcal{Y}}{\partial \mu} = \frac{\partial \tilde{y}}{\partial x}$ for all $F = (\mu, \sigma^2)$ is that*

$$y(x, a) = \gamma x^2 + \Psi(a)x + \Phi(a) \tag{5}$$

for some $\gamma \in \mathbb{R}$ and some C^1 functions Ψ and Φ. In that case, $\mathcal{Y}(\mu, \sigma^2) = Y(\mu) + \gamma\sigma^2$, that is, the long-run expected outcome function is a vertical translation of the long-run outcome response function.

Proof: We have already argued sufficiency. To show necessity, consider the class of probability distributions $G_{x,x_0,p} = (x, x_0 + x; 1 - p, p)$ that take on the value x with probability $1-p$ and the value x_0+x with probability p, and $0 \leq p \leq 1$. The expectation is $\mu = px_0 + x$, while the variance is $\sigma^2 = x_0^2 p(1-p)$. Therefore, if one holds x_0 and p constant, then a change in the mean μ, keeping the variance constant, is equivalent to a change in the value of x. Applying condition (4) to this class of distributions, we get

$$\begin{aligned} \Gamma'(px_0 + x) &= \left.\frac{\partial}{\partial x}\right|_{p,x_0} [(1-p)\Gamma(x) + p\Gamma(x_0 + x)] \\ &= (1-p)\Gamma'(x) + p\Gamma'(x_0 + x). \end{aligned}$$

If condition (4) is to hold for all distributions, it must hold for all values of the distribution parameters x_0 and p within the $G_{x,x_0,p}$ class. That is, we must have

$$\forall x_0,\ \forall x,\ \forall p \in [0,1] \quad \Gamma'(px_0 + x) = (1-p)\Gamma'(x) + p\Gamma'(x_0 + x),$$

that is,

$$\forall x_0,\ \forall x,\ \forall p \in [0,1] \quad \Gamma'\left((1-p)x + p(x_0 + x)\right) = (1-p)\Gamma'(x) + p\Gamma'(x_0 + x),$$

that is

$$\forall x,\ \forall y,\ \forall p \in [0,1] \quad \Gamma'\left((1-p)x + py\right) = (1-p)\Gamma'(x) + p\Gamma'(y).$$

This last property implies that Γ' must be affine (as a function that is both concave and convex), and thus Γ itself must be quadratic. Given that the structure in (3) already allows for an interaction between x and an unspecified function of a (which may include a constant), the restriction of $\Gamma(x)$ to a single quadratic term does not constrain

$3a'\mu^2 + 2b'\mu + c'$.

the model specification beyond what has been shown to be necessary. The last part of the proposition is obtained by specializing the definitions of Y and $\mathcal{Y}$ to the assumed structure for $y(x, a)$. Q.E.D.

Proposition 2 provides an interesting result in light of the common use of the quadratic functional form to represent the relationship between weather and outcomes in empirical work. However, it should not be construed as implying that estimating a quadratic panel regression of the outcome on weather, even with fixed effects, will trace out the long-run expected outcome response function, or even the long-run outcome response function. Mérel and Gammans (2021) demonstrate how a quadratic weather panel leads to biased counterfactual estimates when the structural outcome function is of the form $y(x, a) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3(x - a)^2$, with $\beta_3 < 0$. This is a special case of Equation (5) with $\gamma = \beta_2 + \beta_3$, $\Psi(a) = \beta_1 - 2\beta_3 a$, and $\Phi(a) = \beta_0 + \beta_3 a^2$. The authors show that estimates of the parameters (β_1, β_2) obtained from a simple quadratic panel with fixed effects are biased, affecting identification of both the long-run outcome response function $Y(\mu)$ and the long-run expected outcome response function $\mathcal{Y}(F)$. The intuition behind this result is that although fixed effects account for the function $\Phi(a)$, structurally the slope term $\Psi(a)x$ still depends on actions and thus climate, whereas the simple quadratic specification imposes common coefficients on x across panels.

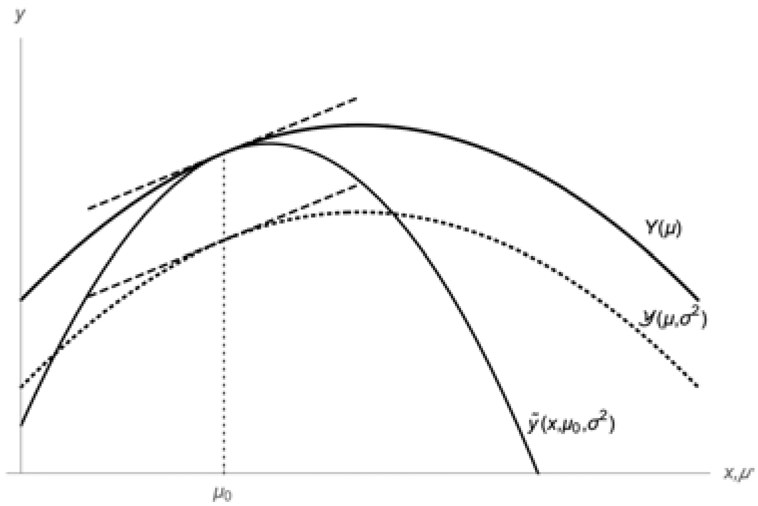
An econometric panel implementation of the structural relationship in Equation (5) that remained totally agnostic about the functions Ψ and Φ would be:

$$y_{it} = \alpha_i + \beta_{1i} x_{it} + \beta_2 x_{it}^2 + \epsilon_{it}.$$

Assuming stationary climates, this regression would identify the function Γ through the β_2 coefficient, but it would not identify $y(x, a)$ itself because the actions a are unobserved and instead left implicit in the coefficients α_i and β_{1i}. Thus, the estimated relationship would be useless to recover the long-run outcome function (or the long-run expected outcome function), and so counterfactual analysis of climate impacts would be precluded.

Finally, note that although Proposition 2 establishes a result about equality of slopes, it does not ensure tangency, as there is no guarantee that $\mathcal{Y}(\mu, \sigma^2) = \tilde{y}(\mu, \mu, \sigma^2)$. Of course, equality in levels is not required for counterfactual analysis of changes in climate, so one should not view this caveat as a weakness. For the sake of completeness, the proposition below establishes that tangency requires $\Gamma(x)$ to be affine, that is, $\gamma = 0$ in Equation (5).

Figure 1 Key outcome functions and their relative positions



Proposition 3 *A necessary and sufficient condition for* $\mathcal{Y}(\mu, \sigma^2)$ *to be tangent to* $\tilde{y}(x, \mu, \sigma^2)$ *at* $x = \mu$ *for all* (μ, σ^2) *is that*

$$y(x, a) = \Psi(a)x + \Phi(a) \tag{6}$$

for some C^1 *functions* Ψ *and* Φ*.*

Proof: Under the structure of Proposition 2, $Y(\mu)$ is tangent to $\tilde{y}(x, \mu, \sigma^2)$ at $x = \mu$ and $\frac{\partial \mathcal{Y}}{\partial \mu}(\mu, \sigma^2) = \frac{\partial \tilde{y}}{\partial x}(\mu, \mu, \sigma^2)$. Since $\mathcal{Y}(\mu, \sigma^2) = Y(\mu) + \gamma\sigma^2$, a necessary and sufficient condition for tangency between $\mathcal{Y}$ and $\tilde{y}$ is that $\gamma = 0$.

2.6 Illustration

Figure 1 depicts the short-run reduced-form outcome function, the long-run outcome response function, and the long-run expected outcome response function in the special case where the structural outcome function has the form $y(x, a) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3(x - a)^2$, with $\beta_3 < 0$ and $\beta_2 < 0$. This structure satisfies Assumption 1 because $\frac{\partial y}{\partial x} = \beta_1 - 2\beta_3 a + 2(\beta_2 + \beta_3)x$ and $\beta_3 \neq 0$. It is a special case of the structure in Proposition 2 with $\gamma = \beta_2 + \beta_3 < 0$, $\Psi(a) = \beta_1 - 2\beta_3 a$, and $\Phi(a) = \beta_0 + \beta_3 a^2$.

Because the restriction of Proposition 1 is satisfied, there is tangency between the short-run reduced-form outcome function corresponding to climate μ_0 (itself equal to the conditional reduced-form outcome function) and the long-run outcome response function at μ_0. Because the restriction of Proposition 2 is satisfied, there is also equality of slopes between the short-run reduced-form outcome function and the long-run expected outcome function at μ_0. But because the restriction of Proposition 3 does not hold, these last two functions are not tangent to each other at μ_0.

3 Consequences for empirical work on optimized outcomes

A large literature has sought to exploit random weather fluctuations to identify the response of optimized or quasi-optimized outcomes to climate (Dell et al., 2014; Blanc and Schlenker, 2017; Auffhammer, 2018). While some of this literature, notably Deschênes and Greenstone (2007), acknowledges at the outset that the identified effect may not reflect long-run adaptation to climate, several papers have invoked the Envelope Theorem to argue that this effect is still relevant for climate change analysis, at least at the margin. A recent working paper goes further by using the Envelope Theorem as a justification for non-marginal analysis, arguing that a global nonlinear relationship between an optimized outcome (GDP) and weather identified in a panel with fixed effects represents a long-run relationship inclusive of climatic adaptation (Deryugina and Hsiang, 2017).

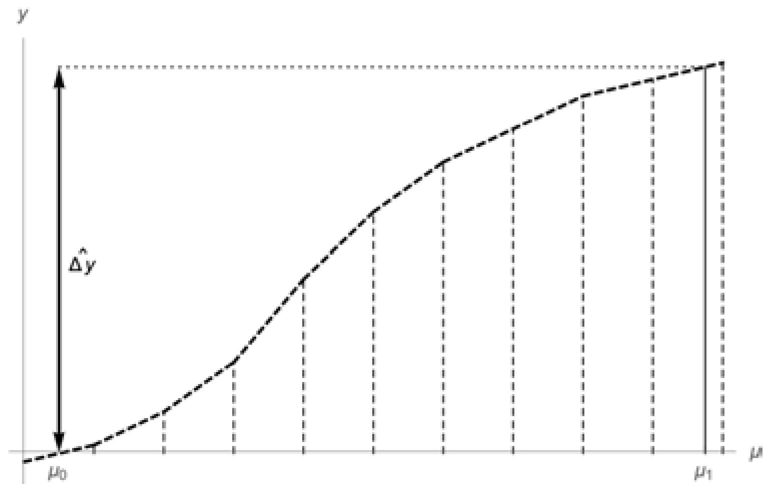
Such a strong result requires more than the Envelope Theorem to hold (Mérel and Gammans, 2021). Nonetheless, it remains true that, to the extent that an Envelope result holds locally, as in the instances highlighted in Section 2, a global long-run relationship may be recovered by stepwise integration, as initially suggested by Hsiang (2016). This is the approach taken in the rest of this paper. While we cannot claim as our own the idea of integrating marginal effects estimated using local time-series variation, to the extent of our knowledge we are the first to implement it empirically.

3.1 Recovering a long-run response by integrating marginal effects across climates

By assuming that the long-run and short-run responses of the outcome are tangent to each other at the mean weather (which, as shown above, is tantamount to assuming the structure in Equation (3) for some unspecified functions Γ, Φ, and Ψ), we can use local time-series variation in weather to identify local marginal responses in a first step, and in a second step integrate stepwise across climates to recover the global relationship between climate and the outcome.

More specifically, assume that we have panel data, where each panel represents a location, e.g., a county. We may estimate marginal responses for each county, or more

Figure 2 Stepwise integration between two climates



Note: Vertical dashed lines delimit climatic intervals. $\hat{\Delta y}$ represents our estimate of the change in outcome resulting from a movement from climate μ_0 to climate μ_1 along the estimated long-run response function.

parsimoniously we may group counties by climates by defining climatic intervals, and estimate local marginal responses for each climate interval. The first step of our analysis thus consists of estimating the following regression:

$$y_{it} = \beta_{\mu(i)} x_{it} + f_{\text{state}(i)}(t) + \alpha_i + \alpha_t + \epsilon_{it} \tag{7}$$

where i denotes a county, t denotes a year, y_{it} denotes the outcome (in our application, agricultural GDP), x_{it} denotes weather, $\mu(i)$ denotes the climatic interval where county i is located, β_μ is the weather slope relevant for climate interval μ, state(i) is the state of county i and $f_{\text{state}(i)}(t)$ is a state-specific time trend, and α_i and α_t are respectively county and year fixed effects.

In the second step, stepwise integration is performed by multiplying local marginal effects by the climatic width of each interval and summing up across intervals located between the initial and final climate values. This allows us to recover the long-run difference in outcomes between an initial climate μ_0 and a final climate μ_1. This procedure is represented graphically on Figure 2 for the case of unidimensional climate.

Contrary to previous work, the approach described above is essentially model-free, or non-parametric, once the choice of relevant weather variables has been made. The global response to climate is identified flexibly by the simple arrangement of short-run slopes across a large number of climates (in our case, 100 intervals for each climatic variable). The only assumptions required for the resulting relationship to identify the underlying long-run response are thus (i) the existence of a common (or global) long-run response across climates, and (ii) the local tangency result. Note that climatic variation is not in any way absent from recovery of the long-run response: although only weather fluctuations are used to identify local marginal effects, the arrangement of these slopes into a global response requires knowledge of mean weather, i.e., climates.

3.2 Addressing omitted variable bias in local slope estimates

While the integration of local slopes along a climate gradient yields a long-run response function under the functional assumptions discussed in Section 2, there remains a possibility that these local slopes, which reflect a marginal response at any given climatic value, may be biased by the presence of omitted variables. Specifically, if there exist time-invariant unobserved factors common to counties in a given climatic interval that contribute to the marginal weather response, i.e., that interact with weather, they will confound the estimated weather slope (to the extent that they are not caused by climate, of course). That is, the weather slope estimated at a given climate may reflect the effect of these factors, in addition to that of climate. This is a problem for counterfactual climate analysis because as counties' climates change under warming, making them cross climatic intervals, these other factors do not change. In order to address this type of bias, we exploit the fact that climates may already have changed during the period of observation, which in our setting spans 17 years. Specifically, in our most demanding specification we include state-level weather slopes as additional covariates, which non-parametrically control for time-invariant, state-specific factors that may interact with weather such as macroeconomic factors and, to some extent, geographic variables such as soils that exhibit some form of spatial continuity. That is, in the first step we estimate the following augmented regression:

$$y_{it} = \beta_{\mu(i,t)} x_{it} + \gamma_{\text{state}(i)} x_{it} + f_{\text{state}(i)}(t) + \alpha_i + \alpha_t + \epsilon_{it} \tag{8}$$

where $\mu(i, t)$ now indicates the climatic interval where county i is present in year t. This latter determination is based on a rolling climate μ_{it} calculated using the weather average over the previous thirty years, $\mu_{it} = \frac{\sum_{s=t-30}^{t-1} x_{is}}{30}$. Due to perfect multicollinearity, we omit the term $\gamma_{\text{state}(i)} x_{it}$ for exactly one state. This normalization implies that the estimated marginal slopes $\hat{\beta}_{\mu}$ represent those relevant for the omitted state. Marginal slopes for counties located in other states are obtained by adding the constant slopes $\gamma_{\text{state}(i)}$ to the set of slopes β_{μ}, yielding state-specific climatic trajectories that are purged of state-level confounders. The inclusion of state-weather interactions as controls implies that identification of weather slopes within a given climate interval (β_μ) now relies on the movement of counties across different climatic intervals over time, as well as within-state climatic variation across counties. To see why, imagine that all counties have a constant climate, and all counties in a state have the same climate. Then, the effect β_μ cannot be identified because all variation in the marginal responses to weather across counties is captured by the state-weather interaction.

4 Empirical implementation

Using the method described in Section 3, we now derive the long-run response of US county-level agricultural GDP to temperature and precipitation.

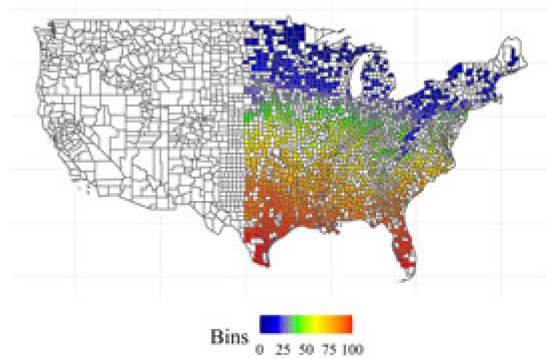
4.1 Data

US county-level data on agricultural GDP over the period 2001–2017 come from the Bureau of Economic Analysis of the US Department of Commerce. As in Schlenker et al. (2006) and Schlenker and Roberts (2009), our analysis focuses on counties east of the 100th meridian, where agriculture is largely non-irrigated. We also remove counties with more than 10% missing values, which leaves us with an almost balanced panel of 1,308 counties representing 71.5% of total US agricultural GDP over the period of investigation. We use historical weather data derived from the PRISM monitoring network (PRISM Climate Group, 2018). The raw data is available at a 4 km resolution and includes daily information on maximum temperature, minimum temperature, and precipitation. Following the approach of Schlenker and Roberts (2009), we aggregate weather data to the county level using farmland areas as weights. We focus on weather and climate during the growing season defined as April to October, and consider two variables: average temperature and total precipitation.

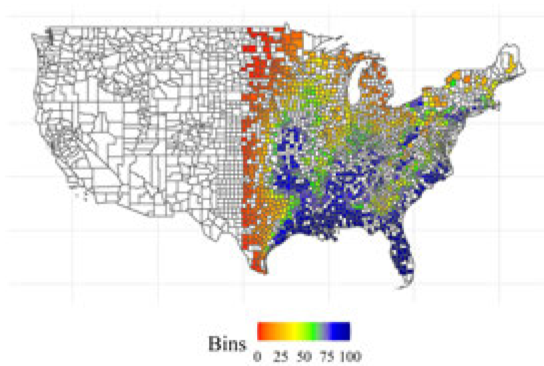
For each county, we calculate the 30-year rolling average of past weather in each year of our sample and use the average of this time series as the central climate value. For each of our two climatic variables, we divide the spectrum of observed central climates into 100 climatic intervals (or bins), so that roughly 13 counties are present in each bin. This choice implies that bins have varying widths; typically those located at the endpoints of the climatic spectrum are wider due to the smaller number of counties with extreme climates. US agriculture spans a wide swath of climates. Growing-season average temperature ranges from 10°C to 29°C. For growing-season precipitation, most

Figure 3 Geographical distribution of central climates

(a) Temperature



(b) Precipitation



of the mass of the distribution is centered around 700 mm, although our sample includes both dramatically drier (<400 mm) and wetter (>1,000 mm) climates. This cross-sectional variation is useful in constructing a long-run response curve whose domain includes most of the future climates likely to be seen as a result of climate change. Figure 3 depicts the geographical distribution of the central climate intervals for counties included in our sample.

4.2 Models

We implement the envelope-gradient-theorem approach on three variants and compare the results to those obtained from a traditional panel approach. Our benchmark panel model is quadratic in temperature and precipitation. As argued in Mérel and Gammans (2021), the quadratic model has been used extensively in the prior climate impact assessment literature. All our estimated models, including the benchmark, include county and year fixed effects as well as state-specific linear time trends to control for variations in the unobserved determinants of agricultural GDP.

Our first envelope-gradient-theorem approach assumes that climates are stationary in each county. County climates are computed using the simple average of climate values over the 17 observation years as $\bar{\mu}_i = \frac{\sum_{t=2001}^{2017} \mu_{it}}{17}$. As a consequence, each county remains in the same climate bin across the entire period of observation. Marginal weather effects are computed by estimating Equation (7). Climate-specific marginal effects are given by the set of estimated coefficients $\hat{\beta}_\mu$.

One shortcoming of the previous approach is that it assumes away climatic trends during the observation period. To the extent that counties have experienced trends in climate, such variation could be exploited to identify the climate-outcome relationship. In fact, variation arising from climatic trends could be viewed as more legitimate than pure cross-sectional climatic variation because it is presumably less prone to confounding factors (Burke and Emerick, 2016). In a second variant, we thus replace the time-invariant central county climates $\bar{\mu}_i$ by their value at any given point in time, μ_{it}. Therefore, counties are allowed to cross climatic intervals based on variation in their rolling climatic average. The estimating equation becomes:

$$y_{it} = \beta_{\mu(i,t)} x_{it} + f_{\text{state}(i)}(t) + \alpha_i + \alpha_t + \epsilon_{it}. \tag{9}$$

This equation is nearly identical to to Equation (7), with the key difference that $\beta_{\mu(i)}$ is replaced with $\beta_{\mu(i,t)}$.

A possible improvement over the approach taken in Equation (9) is to further control for time-invariant factors that may affect the response of the outcome to weather variables. This possibility is explored in our third approach, which relies on estimates of the marginal weather effects obtained from Equation (8). The marginal effect of weather at given climate β_μ is then identified from within-county comparisons across climates visited over time and, because marginal weather effects are allowed to vary systematically across states through the inclusion of state-weather controls, from within-state comparisons across counties.

Like in the previous approach, this strategy relies on temporal variation in climate to identify the marginal effect of climate, and more so since part of the cross-sectional variation in marginal effects has been removed from identification. Importantly, climatic trends tend to be quite small relative to the large cross-sectional climatic variation in our sample. For example, the standard deviation of the stationary climates $\bar{\mu}_i$ is 3.4°C for temperature and 112 mm for precipitation, while the average of the within-county standard deviations of μ_{it} is 0.08°C for temperature and 10.3 mm for precipitation. Despite their small size relative to the cross-sectional variation, these climate trends are statistically significant in many locations, as evidenced in prior work by Burke and Emerick (2016) and Cui (2020). Nonetheless, when implementing this approach we verify that the variation induced by counties crossing climatic intervals is driven at least in part by climate trends, rather than movements in the rolling average driven by random weather shocks. To that effect, we compare the number of interval crossings by a county to the number of intervals "visited." For temperature (resp. precipitation) the median county experiences 3 (resp. 12) climate intervals crossings and visits 3 (resp. 9) climate intervals. On average, 66% (resp. 64%) of crossings are to a previously unvisited climate, suggesting that a meaningful share of these crossings are driven by climate trends, rather than weather fluctuations. Also note that climate trends may be non-monotonic, with periods of cooling followed by periods of warming for example, which implies that re-visits of a climatic interval by a county may be due to "genuine" climatic variation.

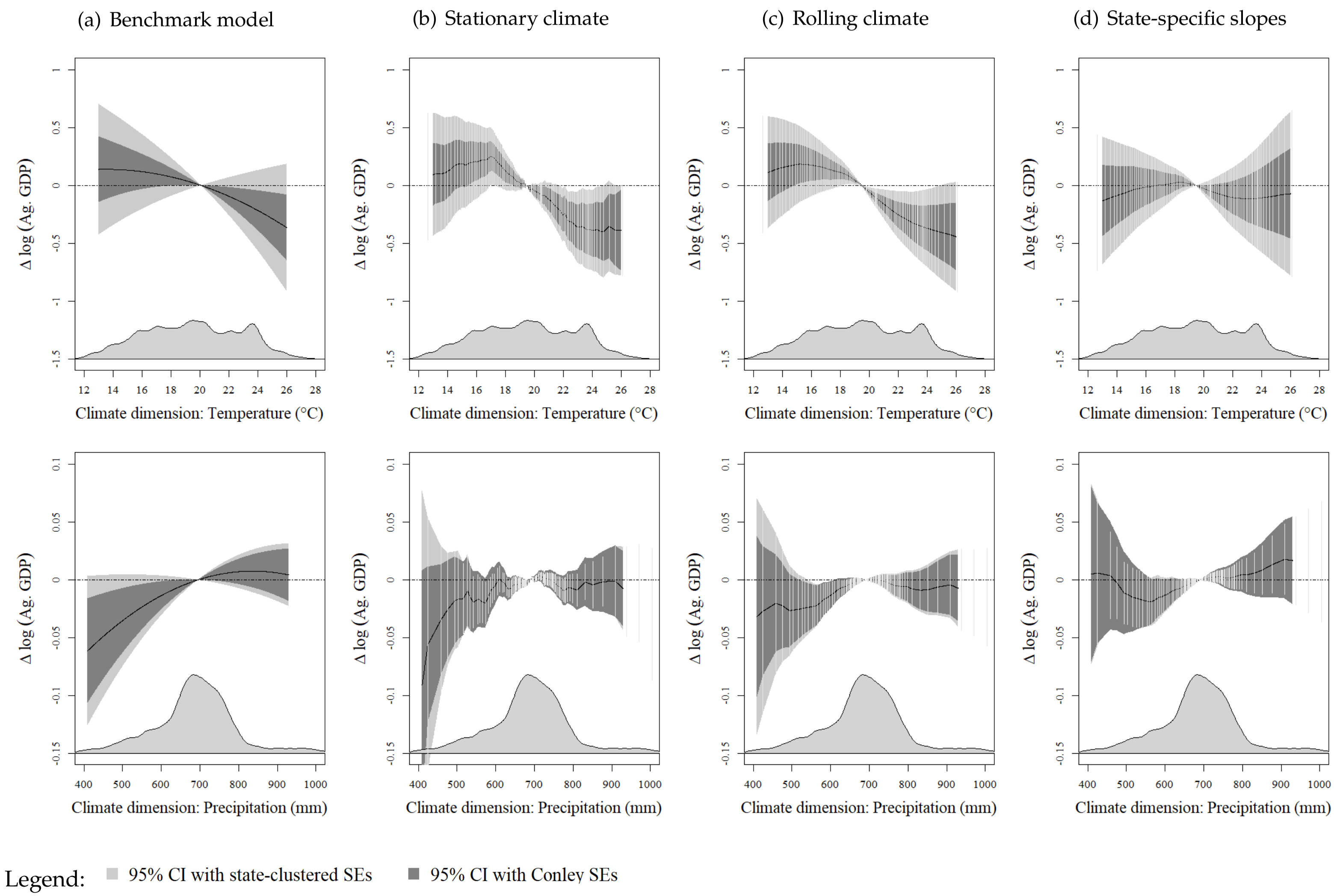
For statistical inference, we rely on a Conley variance-covariance matrix that allows for spatial correlation across adjacent counties, as in Schlenker and Roberts (2009) or Schlenker (2017).[12] For comparison purposes we also show state-clustered standard errors. We use our estimate of the variance-covariance matrix of the $\hat{\beta}_\mu$ coefficients to construct confidence intervals around the long-run response functions of agricultural GDP to each climatic variable. These confidence intervals demonstrate the uncertainty pertaining to a counterfactual change from the median of the county-specific climates $\bar{\mu}_i$ to any climatic value, lower or higher. They thus answer the question "how confident are we that a change away from the median climate would lead to a difference in agricultural GDP?" Since predicted changes in the outcome are a linear combination of the estimated $\hat{\beta}_\mu$ parameters, the confidence intervals can be easily constructed using the variance-covariance matrix of coefficient estimates.

4.3 Long-run responses to climate

Figure 4 depicts the long-run response curves for temperature and precipitation, for the benchmark panel model and each of the three envelope-gradient-theorem models. In columns (b), (c) and (d), grey vertical lines delineate the 100 bins used in the estimation. The distribution of climates is shown at the bottom of each panel and

[12] We allow for spatial correlation across neighboring counties using a neighboring lag structure, rather than geographical distance. Correlation is allowed up to the third-degree neighbor.

Figure 4 Long-run responses to climate



does not depend on the model. Confidence intervals based on Conley standard errors and more conservative state-clustered standard errors are shown with dark and light shading, respectively.

Column (a) of Figure 4 shows long-run responses for the benchmark panel model. This model predicts significantly negative effects of increasing climate temperature above the median value when using Conley standard errors. The estimated precipitation response is an inverted U-shape, with the response peaking around 800 mm of cumulative precipitation. Thus, for most climates agricultural GDP is predicted to increase with additional precipitation according to this model. Note, however, that this precipitation response is one order of magnitude smaller than the response to temperature. This finding is congruent with previous evidence on US agricultural yields, for instance Schlenker and Roberts (2009).

The patterns depicted in columns (b) and (c) of Figure 4 are reassuring about the ability of the envelope-gradient-theorem approach to deliver meaningful and useful response functions. First, despite the demands put on our data (marginal effects are only identified from a few counties in each climatic interval), the counterfactual effects we derive are relatively precisely estimated, with confidence intervals comparable in size to those obtained in the parametric benchmark panel. Second, these effects largely conform with expectations grounded in previous parametric work on agricultural yields. Notably, the marginal effect of warming is significantly negative. Interestingly, we find a more negative marginal impact in the middle of the distribution of temperatures than at the upper end. One explanation could be that once crop yields have suffered from heat, the marginal effect of additional heat diminishes. Intuitively, in the limit where the yield reaches zero, there must be a tapering of the marginal effect. Since nothing in our two-step procedure forces this change in curvature to arise, these results suggest that empirical analyses of temperature and agricultural outcomes ought to allow for varying convexity in the temperature response. In contrast to the benchmark panel, which showed a positive response of agricultural GDP to precipitation over most of the climate distribution, the precipitation responses shown in columns (b) and (c) are less clear-cut. However, their order of magnitude is preserved.

Overall, results from the first two variants of the envelope-gradient-theorem approach are comparable, even if the second one produces smoother curves. This finding could in fact be expected, as unlike column (b), column (c) exploits the movement of counties across neighboring climate intervals over time for identification. As a result, time-invariant, county-level idiosyncratic factors that may affect marginal responses to weather tend to be captured in adjacent intervals, as opposed to a single interval, resulting in a smoothing of effects.

Column (d) of Figure 4 depicts the long-run responses to each climatic variable in our most flexible variant, which more heavily relies on climatic trends for the identification of climate-specific marginal effects. In this model, counties' climatic response functions are allowed to vary systematically by state. Namely, state-specific trajectories differ from one another by a constant slope adjustment, added to each of the 100 climatic intervals. The figure represents the long-run response of a fictitious county endowed with the simple average of counties' slope parameters.[13] Because all model parameters are estimated within the same regression, the confidence intervals shown in column (d) of Figure 4 also account for the covariance between the $\gamma_{\text{state}(i)}$ parameters. The average response functions obtained in this model are not statistically significant, perhaps reflecting the limitations of the envelope-gradient theorem framework in accommodating additional controls that interact with weather, but also likely due to the limited amount of "useful" climatic variation left to identify the β_μ parameters once state-weather interactions have been introduced as covariates, as climatic variation over time conditional on location is limited (see Section 4.2), and so is climatic variation within states. Fisher et al. (2012) make a comparable remark in the context of the parametric panel approach regarding the use of state-by-year fixed effects in regressions of county agricultural profits on weather, as these fixed effects absorb a large amount of useful weather variation, leading to statistically insignificant estimates.

4.4 Climate counterfactuals

We use our calculated long-run response functions to derive the county-level impacts of a 2°C uniform warming on agricultural GDP. These impacts compare predicted values under the 2001–2017 climate to a counterfactual where the climate is 2°C warmer for all counties. At the county level, the impact on the log agricultural GDP is given by the linear combination $\sum_{b\in B}\hat{\beta}_b\Delta\text{Temp}_{b,i}$ where B is the set of temperature bins, $\hat{\beta}_b$ is the estimated slope of bin b, and $\Delta\text{Temp}_{b,i}$ is the length of the temperature interval that county i crosses in bin b under a 2°C warming scenario.[14] The county-level impact on agricultural GDP is thus $\widehat{\Delta}_i = \exp\left(\sum_{b\in B}\hat{\beta}_b\Delta\text{Temp}_{b,i}\right) - 1$, expressed in percentage terms. We then aggregate these impacts at the country level by taking the weighted

[13]There are as many slopes as states in the sample, but we average over counties to capture the fact that some states have more counties than others.

[14]If county i does not cross bin b, the length is zero; if it fully crosses it the length is that of the bin itself. The only bins partially crossed are those located at the endpoints of the county's climatic trajectory. See Figure 2.

average $\sum_{i \in C} w_i \widehat{\Delta}_i$, where C denotes the set of counties included in the simulation and w_i is the share of the agricultural GDP of county i in the total agricultural GDP of the included counties.[15]

In this simulation, we do not include the warmest counties for which counterfactual temperatures are projected to lie too far away in the warmest bin.[16] The rationale is that each estimated slope is only valid locally. We cannot evaluate the damage for the warmest climates without assuming a constant slope after a certain threshold, an assumption we do not wish to make. However, the convexity of the long-run response functions estimated using our approach suggests that the damage would be lower than average for these warmest counties.

Figure 5 shows county-level warming impacts for the benchmark panel model and the three envelope-gradient-theorem models. The x-axis represents the county's reference period climate temperature value and the y-axis represents the change in the logarithm of agricultural GDP associated with a 2°C warming. For each county-level impact, we show a 95% confidence interval based on the Conley standard errors. In each panel, we further report cumulative US impacts.

Panel (a) of Figure 5 shows impacts for the benchmark panel model. This model predicts that currently warm counties are harmed by increases in temperature, and significantly so. Predicted impacts range from near zero for counties near the median climate to losses of over 10% in the warmest counties. Overall, we find a statistically and economically significant US-wide agricultural GDP loss of 7.83%.

Panels (b) through (d) of Figure 5 show impacts for the envelope-gradient-theorem models. Panel (b) provides results for the model that uses a stationary definition of climate. This model implies much more detrimental effects than the benchmark panel model in moderate climates, but less severe decreases in warmer climates. This discrepancy is due to the convex shape of the estimated long-run response to temperature. Aggregating across counties, we find that a 2°C warming is associated with a 9.45% agricultural GDP loss, a decline that is both economically and statistically significant, and slightly greater in magnitude than the aggregate loss implied by the benchmark panel model. The results from the variant that uses a rolling climate definition without allowing for state-specific slopes are shown in panel (c). Overall, county-level and

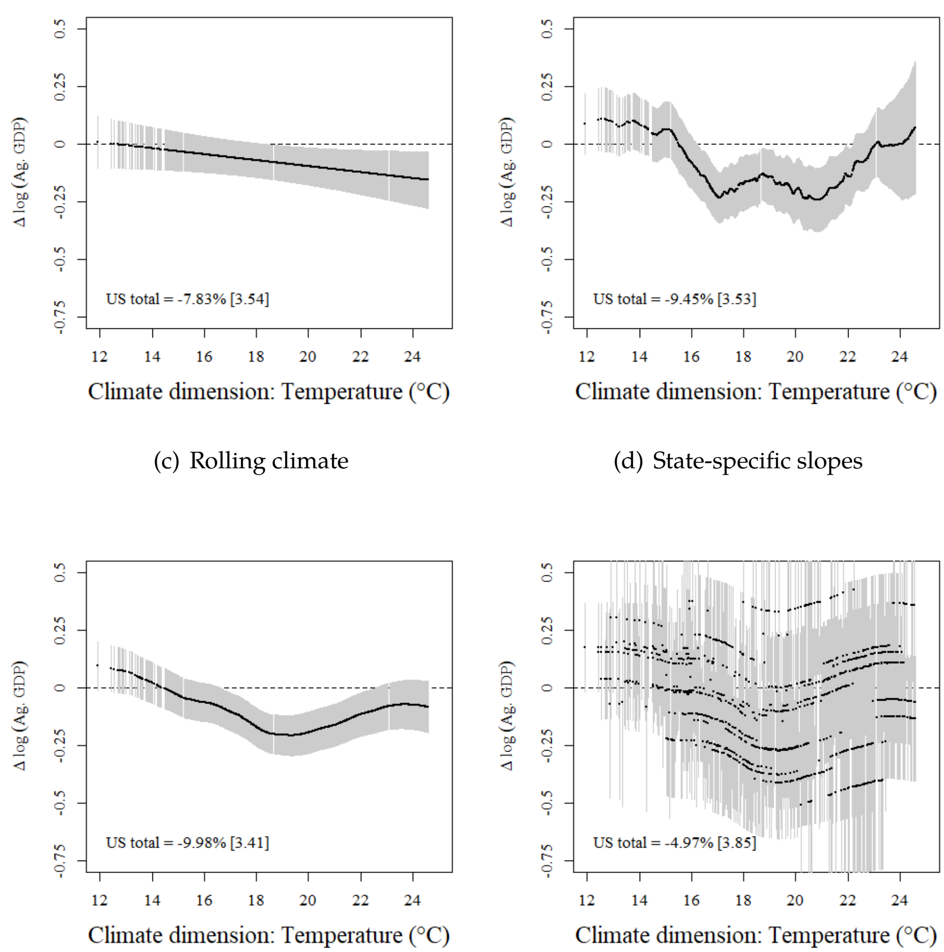
[15] To obtain w_i, we first compute the county-specific average of the agricultural GDP over the 17 years in our sample, say agGDP_i, and then construct the ratio $w_i = \frac{\text{agGDP}_i}{\sum_{j \in C} \text{agGDP}_j}$.

[16] These counties have climates that are warmer than the lower bound of the last bin (26.07°C) plus the maximum width of the middle bins (0.52°C) minus 2°C. There are 78 such counties out of the 1,308 used in the estimation (6.0%). They produced 6.6% of the US total agricultural GDP on average over 2001–2017.

Figure 5 Simulated impact of a +2°C scenario on county agricultural GDP

(a) Benchmark model (b) Stationary climate

(c) Rolling climate (d) State-specific slopes



Legend: Each dot corresponds to a county. The grey segments represent 95% confidence intervals using Conley standard errors. In each panel, the estimate of the total effect on the US national average agricultural GDP is given together with its Conley standard error in brackets.

aggregate results are comparable to those from the stationary climate variant. In line with our estimated long-run response functions, the distribution of impacts across climates appears smoother. In this variant, we find a statistically significant aggregate agricultural GDP loss of 9.98%.

We provide results for our most flexible model, which allows for state-specific slope coefficients, in panel (d) of Figure 5. In this variant, slope estimates are purged of state-specific time-invariant factors that interact with weather. Consistent with the climatic response functions depicted in Figure 4, this innovation leads to substantially less precise impact estimates. Additionally, since counties are assigned a state-level weather slope coefficient, there is substantial variation in the estimated impact of warming across counties, even if they share similar climates, leading to large discontinuities in the distribution of effects across the climate distribution, and to effects that appear to be located on parallel curves. The points located on each of these curves correspond to counties in the same state or in states with similar slopes. In this very demanding estimation framework, increases in temperature are associated with agricultural GDP losses for the majority of counties (60%). The aggregate damage is negative, although it is less severe than the estimates in the other specifications and not statistically significant at conventional levels. Estimated damages for individual counties also tend to be negative, albeit only significantly so for a minority of them (23%).

5 Conclusion

A full accounting of the consequences of future climate change on outcomes of economic interest remains an important, yet elusive, research goal. There has been a longstanding debate on the ability of methodologies that rely on random weather fluctuations to provide information on the long-run effects of climate change, accounting for adaptation by economic agents (Kolstad and Moore, 2020; Lemoine, 2021; Mérel and Gammans, 2021). This paper provides a set of necessary and sufficient conditions for the short-run response to weather and the long-run response to climate to be tangent to each other at the expected weather irrespective of the weather distribution. Delineating these conditions is important because the tangency property enables the use of weather fluctuations to identify long-run marginal effects. We then introduce an empirical methodology to estimate marginal effects that are specific to a given climatic interval. Using the Gradient Theorem, we demonstrate how these marginal effects can be integrated to construct a common long-run response to climatic variables, as initially proposed by Hsiang (2016).

We apply the envelope-gradient-theorem approach to a panel of weather and agricultural GDP data from the US and compare our results to those obtained from a benchmark quadratic panel model. We find qualitatively similar long-run responses to climate across specifications, except in our most stringent specification. Our results indicate that, despite the considerable demands put on our data, the approach can deliver long-run response functions that are coherent with agronomic expectations and prior parametric work, with some important nuances. For instance, we provide empirical support for changes in convexity in the effects of temperature on agricultural GDP that are not captured in the benchmark panel model.

We then consider the effect of a 2°C uniform warming on US agricultural GDP. Using the long-run climate responses derived from the envelope-gradient-theorem approach, we calculate county-level impacts. We find these impacts to be comparable across models, at least in the aggregate, except for the most demanding specification, which implies an aggregate effect that is smaller in magnitude, and more noisy. Importantly, our approach delivers county-level impact estimates that are different from those of the benchmark panel, especially for counties at either end of the climate distribution. Our specification allowing for a rolling climate definition predicts an aggregate damage to agricultural GDP after adaptation of 10% under the +2°C scenario. This effect is slightly larger in magnitude than the aggregate damage estimate from the benchmark panel (8%).

References

Auffhammer, M. (2018). Quantifying economic damages from climate change. *Journal of Economic Perspectives*, 32(4):33–52.

Blanc, E. and Schlenker, W. (2017). The use of panel models in assessments of climate impacts on agriculture. *Review of Environmental Economics and Policy*, 11(2):258–279.

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

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Kopp, R. E., McCusker, K., Nath, I., Rising, J., Rode, A., Seo, H. K., Viaene, A., Yuan, J., and Zhang, A. T. (2020). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *NBER Working Paper No. 27599*.

Cui, X. (2020). Climate change and adaptation in agriculture: Evidence from us cropping patterns. *Journal of Environmental Economics and Management*, page 102306.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What Do We Learn from the Weather? The New Climate-Economy Literature. *Journal of Economic Literature*, 52(3):740–98.

Deryugina, T. and Hsiang, S. (2017). The Marginal Product of Climate. *NBER Working Paper No. 24072*.

Deschênes, O. and Greenstone, M. (2007). The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather. *The American Economic Review*, 97(1):354–385.

Fisher, A. C., Hanemann, W. M., Roberts, M. J., and Schlenker, W. (2012). The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment. *The American Economic Review*, 102(7):3749–3760.

Guo, C. and Costello, C. (2013). The value of adaption: Climate change and timberland management. *Journal of Environmental Economics and Management*, 65(3):452–468.

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics*, 8:43–75.

Kelly, D. L., Kolstad, C. D., and Mitchell, G. T. (2005). Adjustment costs from environmental change. *Journal of Environmental Economics and Management*, 50(3):468–495.

Kolstad, C. D. and Moore, F. C. (2020). Estimating the economic impacts of climate change using weather observations. *Review of Environmental Economics and Policy*, 14(1):1–24.

Lemoine, D. (2021). Estimating the Consequences of Climate Change from Variation in Weather. *NBER Working Paper No. 25008*.

Maue, C. C. and Kolstad, C. D. (2020). Exact integration of climate and weather in estimating climate damage. Working Paper.

Mendelsohn, R., Nordhaus, W. D., and Shaw, D. (1994). The Impact of Global Warming on Agriculture: A Ricardian Analysis. *The American Economic Review*, 84(4):753–771.

Mérel, P. and Gammans, M. (2021). Climate econometrics: Can the panel approach account for long-run adaptation? *American Journal of Agricultural Economics*, forthcoming.

Nicholson, W. and Snyder, C. M. (2016). *Microeconomic Theory: Basic Principles and Extensions, Twelfth Edition*. Cengage Learning.

Nordhaus, W. D. (2010). The economics of hurricanes and implications of global warming. *Climate Change Economics*, 1(01):1–20.

Ortiz-Bobea, A. (2020). The Role of Nonfarm Influences in Ricardian Estimates of Climate Change Impacts on US Agriculture. *American Journal of Agricultural Economics*, 102(3):934–959.

Perloff, J. M. (2016). *Microeconomics: Theory and Applications with Calculus, Fourth Edition*. Pearson.

PRISM Climate Group (2018). Oregon State University, http://prism.oregonstate.edu, created 8 October 2018.

Schlenker, W. (2017). Inter-annual Weather Variation and Crop Yields. Working Paper, Dept. of Economics and School of International and Public Affairs, Columbia University.

Schlenker, W., Hanemann, W. M., and Fisher, A. C. (2005). Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach. *The American Economic Review*, 95(1):395–406.

Schlenker, W., Hanemann, W. M., and Fisher, A. C. (2006). The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions. *The Review of Economics and Statistics*, 88(1):113–125.

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

Shrader, J. (2021). Improving Climate Damage Estimates by Accounting for Adaptation. Working paper, Columbia University, New York, NY.

Simon, H. A. (1956). Dynamic programming under uncertainty with a quadratic criterion function. *Econometrica*, 24:74–81.

Theil, H. (1954). Econometric Models and Welfare Maximisation. *Weltwirtschaftliches Archiv*, 72:60–83.

Theil, H. (1957). A note on certainty equivalence in dynamic planning. *Econometrica*, 25:346–349.

Appendices

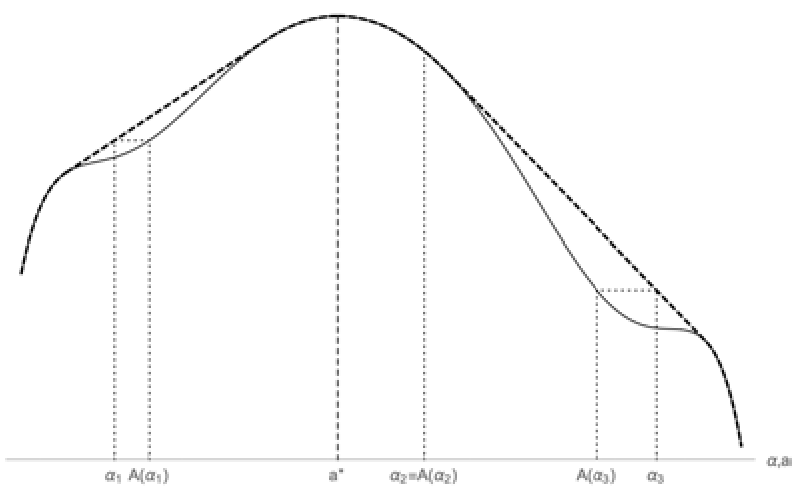
A Concavity of $y(x, a)$ with respect to a

We assume that for all x, $y(x, a) \equiv y_x(a)$ has a unique maximizer characterized by $y'_x(a) = 0$ (that is, there is only one critical point at which the derivative function is zero). Now consider the upper concave envelope of $y_x(\cdot)$, defined as

$$\tilde{y}_x \equiv \inf \{u | u \text{ is convex and } u(a) \geq y_x(a) \; \forall a \in U\}.$$

A representation of such an envelope is depicted in Figure A.1.

Figure A.1 The upper concave envelope



Since y_x has only one critical point, say a^*, at which it is maximized, we must have that $y'_x(a) > 0 \; \forall a < a^*$ and $y'_x(a) < 0 \; \forall a > a^*$. Further denote by y_- the restriction of y to the range $a \leq a^*$ and y_+ its restriction to the range $a > a^*$. Both y_- and y_+ are monotone and therefore bijective. Now consider the following change of variable:

$$a = A(\alpha) = \begin{cases} y_-^{-1}(\tilde{y}(\alpha)) & \text{if } \alpha < a^* \\ y_+^{-1}(\tilde{y}(\alpha)) & \text{if } \alpha > a^* \end{cases}.$$

The change of variable is illustrated in Figure A.1 for various values of α. It is clear that it is well defined because it is both injective (for each α, there is a unique $A(\alpha)$) and surjective (for each a, there is a unique α such that $a = A(\alpha)$). The composite function $\alpha \mapsto y(A(\alpha))$ is by construction equal to the function $\tilde{y}(\alpha)$, therefore it is concave.

B Relationship with Theil (1954)'s model

Slightly adapting his notation to make it more congruent with ours, the structural outcome function considered in Theil (1954) has the following form:

$$W(\mathbf{a},\mathbf{x}) = A(\mathbf{a}) + \sum_{h=1}^{n} A_h(\mathbf{a}) y_h(\mathbf{a},x_h) + \sum_{h=1}^{n}\sum_{k=1}^{n} A_{hk} y_h(\mathbf{a},x_h) y_k(\mathbf{a},x_k) \tag{A-1}$$

where $\mathbf{x}$ and $\mathbf{y} \equiv (y_1,\ldots,y_n)$ are random vectors of the same dimension n and

$$y_h(\mathbf{a},x_h) = f_h(\mathbf{a}) + x_h, \quad h = 1,\ldots,n \tag{A-2}$$

for some unspecified functions f_h. The functions A and A_h are also left unspecified, while the A_{hk} are taken as parametric, that is, they do not depend on $\mathbf{a}$. The vector $\mathbf{a}$ is interpreted as actions (policies) while the vector $\mathbf{x}$ represents random environmental variables (weather in our case) and the vector $\mathbf{y}$ represents "indirect variables" that enter the welfare function and depend both on actions and environmental variables. In an application to farm profits, the indirect variables could be revenue (which depends on weather through yields and on long-run actions such as crop choice) and costs (which may depend on weather through short-run adaptations such as variation in irrigation intensity and on long-run actions such as crop choice). In that sense, the structural outcome function in (A-1) is more "structural" than ours as the random environmental variables enter the objective function through the set of functions $y_h(\mathbf{a},x_h)$ rather than by themselves. Note however that if the functions y_h do not depend on $\mathbf{a}$, then the model is strictly equivalent to ours. Within this structure, Theil (1954) shows that a certainty equivalence principle holds.

We will now show that the restrictions embedded in Equation (A-1) are, in fact, equivalent to those in Equation (3). Plugging the identities in Equation (A-2) into Equation (A-1), we obtain

$$\begin{aligned} W(\mathbf{a},\mathbf{x}) \quad = \quad & A(\mathbf{a}) + \sum_{h=1}^{n} A_h(\mathbf{a}) f_h(\mathbf{a}) + \sum_{h=1}^{n}\sum_{k=1}^{n} A_{hk} f_h(\mathbf{a}) f_k(\mathbf{a}) + \sum_{h=1}^{n} A_h(\mathbf{a}) x_h \\ & + \sum_{h=1}^{n} x_h \sum_{k=1}^{n} (A_{hk} + A_{kh}) f_k(\mathbf{a}) + \sum_{h=1}^{n}\sum_{k=1}^{n} A_{hk} x_h x_k \end{aligned}$$

which can be rewritten as

$$W(\mathbf{a},\mathbf{x}) = B(\mathbf{a}) + \sum_{h=1}^{n} B_h(\mathbf{a}) x_h + \sum_{h=1}^{n}\sum_{k=1}^{n} A_{hk} x_h x_k. \tag{A-3}$$

where $B(\mathbf{a}) \equiv A(\mathbf{a}) + \sum_{h=1}^{n} A_h(\mathbf{a}) f_h(\mathbf{a}) + \sum_{h=1}^{n}\sum_{k=1}^{n} A_{hk} f_h(\mathbf{a}) f_k(\mathbf{a})$ and $B_h(\mathbf{a}) \equiv A_h(\mathbf{a}) + \sum_{k=1}^{n} (A_{hk} + A_{kh}) f_k(\mathbf{a})$.

Apart from the fact that the actions $\mathbf{a}$ and the random variables $\mathbf{x}$ are multivariate, Equation (A-3) has the same structure as Equation (3), where actions are restricted to interact only with a linear function of $\mathbf{x}$, but $\mathbf{a}$ and $\mathbf{x}$ are allowed to affect the outcome in an otherwise unrestricted fashion. (It could appear, in addition, that $\mathbf{x}$ is restricted to enter quadratically into the welfare function. This is not an essential restriction, as one could easily add a function $\Gamma(\mathbf{x})$ to Equation (A-1) without changing the analysis, as there would be no interaction between this new function and the decision variables $\mathbf{a}$.)

The necessity proof in Section 2.3 extends to the case where $\mathbf{x}$ and $\mathbf{a}$ are multivariate in a straightforward fashion. Because the criterion is to ensure that $\tilde{\mathbf{a}}(F) = \hat{\mathbf{a}}\left(\mu(F)\right)$ for *all* (joint) distributions F of the random vector $\mathbf{x}$, one can successively consider degenerate distributions where only one of the random variables in $\mathbf{x}$ is truly random and all others are taken as certain. Each iteration h will lead to the conclusion that the gradient of W with respect to $\mathbf{a}$ must be an affine function of x_h.