Empirical quantification of how the climate impacts socio-economic outcomes

Jonathan Proctor
University of British Columbia

July 16^{th}, 2024

Climate Change AI Virtual Summer School

How will climate change influence human wellbeing?

1. How will greenhouse gas emissions and climate change influence the environment (temperature, rainfall, cloud cover, soil moisture...)?

2. How will changes in the environment influence human wellbeing?

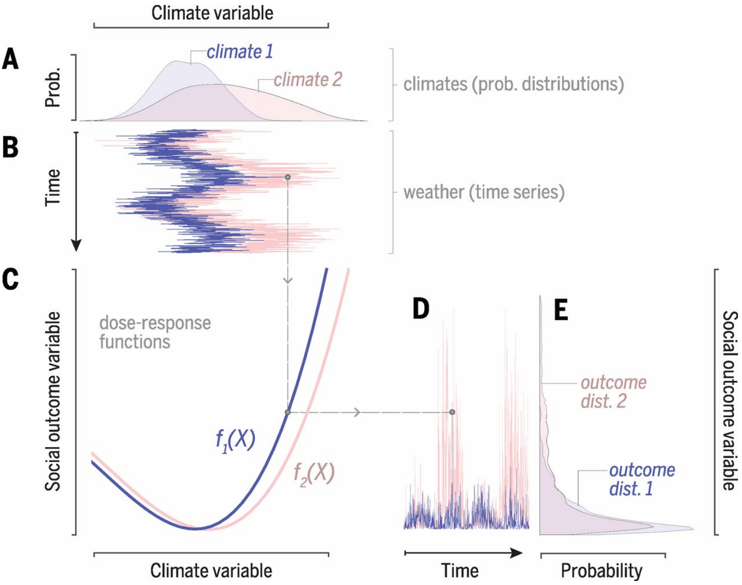
How will climate change influence human wellbeing?

1. How will greenhouse gas emissions and climate change influence the environment (temperature, rainfall, cloud cover, soil moisture...)?
   - Often estimated using earth system models.

2. How will changes in the environment influence human wellbeing?
   - You'll learn how to estimate this today.

$$\frac{\partial \text{Wellbeing}}{\partial \text{CO}_2} = \frac{\partial \text{Wellbeing}}{\partial \text{Temperature}} \frac{\partial \text{Temperature}}{\partial \text{CO}_2} \tag{1}$$

Projecting climate impacts using dose-response functions



(Carleton & Hsiang, 2016)

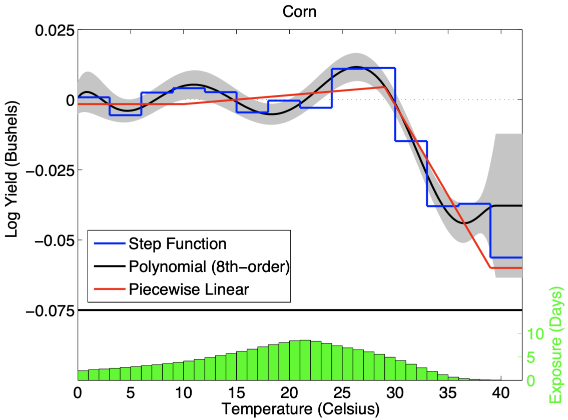
Estimating climate-outcome dose-response functions

We learn the influence of climatic variables on social outcomes by analyzing their historical associations.

- Within a state over time, how do higher than expected mortality rates associate with higher than average temperatures?
- Within a country over time, how do lower than expected crop yields associate with lower than expected soil moisture?

By analyzing many changes in the social outcome variable and climate variable over time we can sketch out the relationship between the two, which we call a "dose-response function".

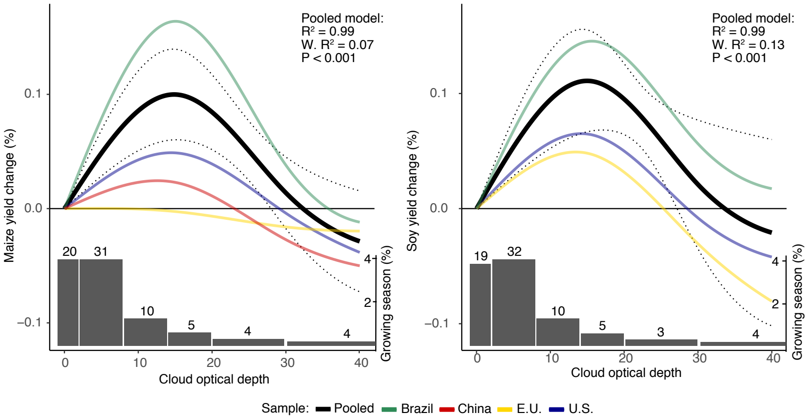
The impact of temperature on crop yields



(Schlenker & Roberts 2009)

“Average yields are predicted to decrease by 30–46% before the end of the century under the slowest warming scenario and decrease by 63–82% under the most rapid warming scenario.”

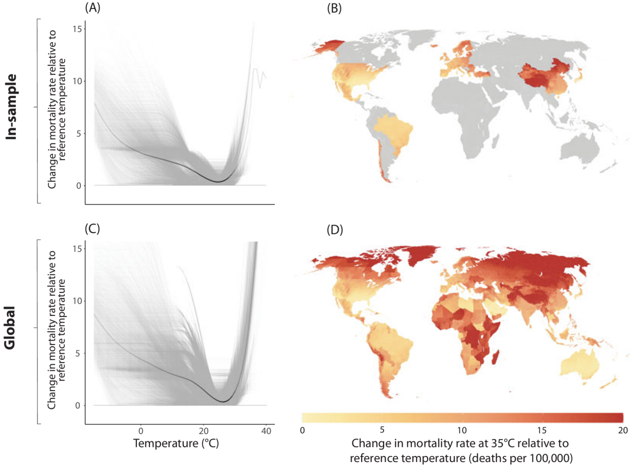
The impact of sunlight on crop yields



(Proctor 2021)

"An additional day of optimal cloud cover, relative to a clear-sky day, increases maize and soy yields by 0.4%."

The impact of temperature on human mortality



(Carleton et al., 2022)

"The release of an additional ton of CO_2 today will cause mortality-related damages of \$36.6 under a high-emissions scenario"

The impact of temperature on human conflict

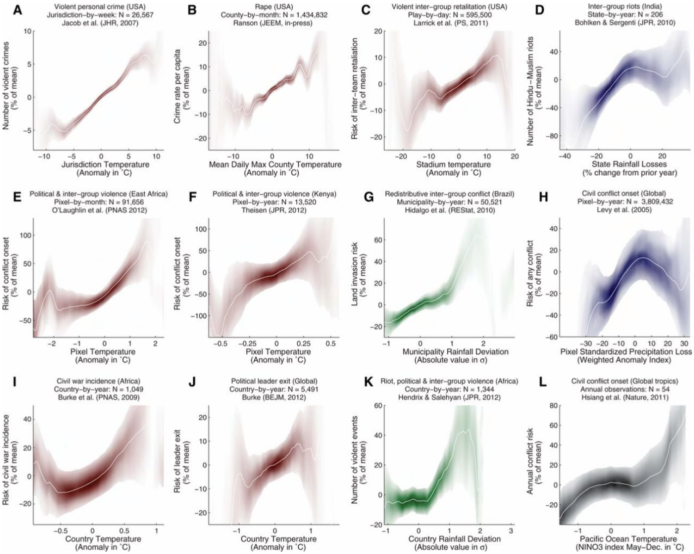
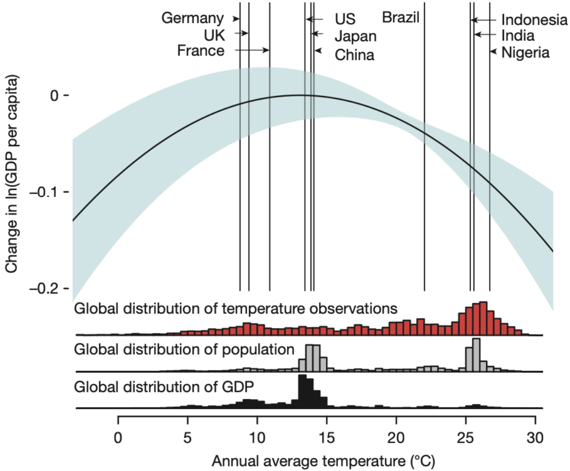


Fig. 2 Empirical studies indicate that climatological variables have a large effect on the risk of violence or instability in the modern world.

(Hsiang, Burke, & Miguel, 2013)

The impact of temperature on GDP

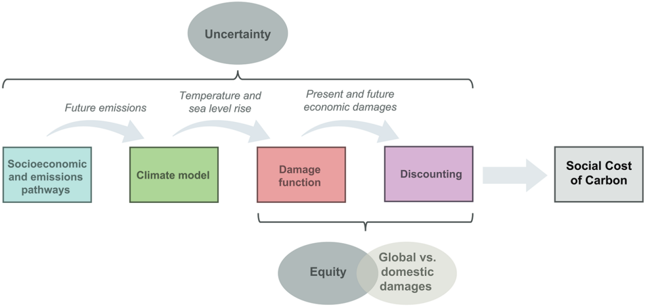


(Burke, Hsiang, & Miguel, 2015)

"Economic productivity is non-linear in temperature for all countries, with productivity peaking at an annual average temperature of 13C and declining strongly at higher temperatures."

Dose-response or "damage" functions provide the foundation for quantifying the "Social Cost of Carbon" (SCC)

- The social cost of carbon is the "monetized value of all future net damages associated with a 1 metric ton increase in CO_2 emissions."

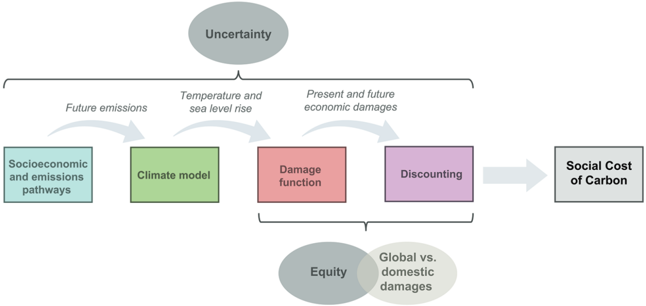


(Carleton & Greenstone, 2022)

- The SCC forms the basis of regulatory cost benefit analysis and carbon policy

Dose-response or "damage" functions provide the foundation for quantifying the "Social Cost of Carbon" (SCC)

- The social cost of carbon is the "monetized value of all future net damages associated with a 1 metric ton increase in CO_2 emissions."



(Carleton & Greenstone, 2022)

- The SCC forms the basis of regulatory cost benefit analysis and carbon policy; yet, many dose-response functions are unknown.

How to estimate climate-outcome dose-response functions

For a social outcome, y, and climate variable(s), $\mathbf{C}$, measured in locations, i, and time periods, t, we can use linear regression to learn the function f() that maps variation in weather onto variation in the social outcome of interest:

$$y_{it} = f(\mathbf{C}_{it}) + \text{controls} + \epsilon_{it}$$

There are two primary challenges when doing so:

1. identifying a causal effect
2. estimating a local non-linear response using aggregated data

Identifying a causal dose-response function

- To accurately predict how a change in climate will influence social outcomes, we need to recover the causal effect of C on y.

- Regressions of y on C could be confounded by variables omitted from the regression that co-vary with both y and C.

Identifying a causal dose-response function

- To accurately predict how a change in climate will influence social outcomes, we need to recover the causal effect of C on y.

- Regressions of y on C could be confounded by variables omitted from the regression that co-vary with both y and C.
  - For example, the northern U.S. is generally both richer and cooler than the southern U.S.; but there are many factors that differ between the two other than climate.

Identifying a causal dose-response function

- To accurately predict how a change in climate will influence social outcomes, we need to recover the causal effect of C on y.

- Regressions of y on C could be confounded by variables omitted from the regression that co-vary with both y and C.
    - For example, the northern U.S. is generally both richer and cooler than the southern U.S.; but there are many factors that differ between the two other than climate.
    - Using spatial differences in temperature and wealth to predict future climates would implicitly assume that the North will become more like the South as the climate warms.

Identifying a causal dose-response function

Ideally, we would have two identical Earths; we could "treat" one Earth by increasing its temperature and compare with the control Earth.

Identifying a causal dose-response function

Ideally, we would have two identical Earths; we could "treat" one Earth by increasing its temperature and compare with the control Earth.

- In practice, a region in one year can act as a "control" for the same region in a different year with a different temperature "treatment"
- Approximates a randomized control trial, since year-to-year weather is as good as randomly assigned w.r.t. most social outcomes.

Implementation using panel fixed-effect regression

- **Panel**: repeated observations of y and C in multiple locations over time. Many time series.
- **Fixed effect**: a set of intercepts, often high dimensional and controlling for time or space; e.g., an intercept for each county.
- **Regression**: ordinary least squares linear regression.

Implementation using panel fixed-effect regression

- **Panel**: repeated observations of y and C in multiple locations over time. Many time series.
- **Fixed effect**: a set of intercepts, often high dimensional and controlling for time or space; e.g., an intercept for each county.
- **Regression**: ordinary least squares linear regression.

Approach: isolate temporal variation within locations over time by including a set of fixed effects λ_i that absorb all variation in y and C across space:

$$y_{it} = f(\mathbf{C}_{it}) + \lambda_i + \epsilon_{it}$$

- Estimates f comparing temporal anomalies in y with anomalies in C.
- Can also include temporal fixed effects – e.g., time trends ($\alpha_i t$).

Practice question

Inclusion of a temporal fixed effect, ϕ_t, (e.g., a fixed effect representing an intercept for each year) in this panel fixed effect regression:

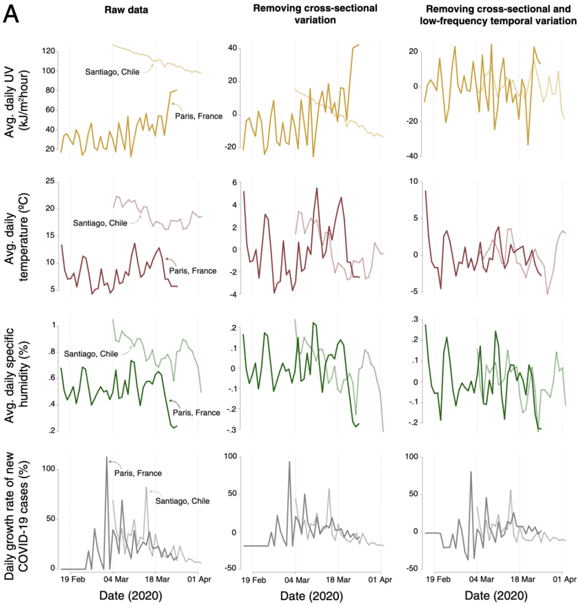
$$y_{it} = f(\mathbf{C}_{it}) + \phi_t + \epsilon_{it}$$

would (choose 1):

1. Effectively demean all observations within each time period t, so that f is estimated using only spatial variation within time periods.

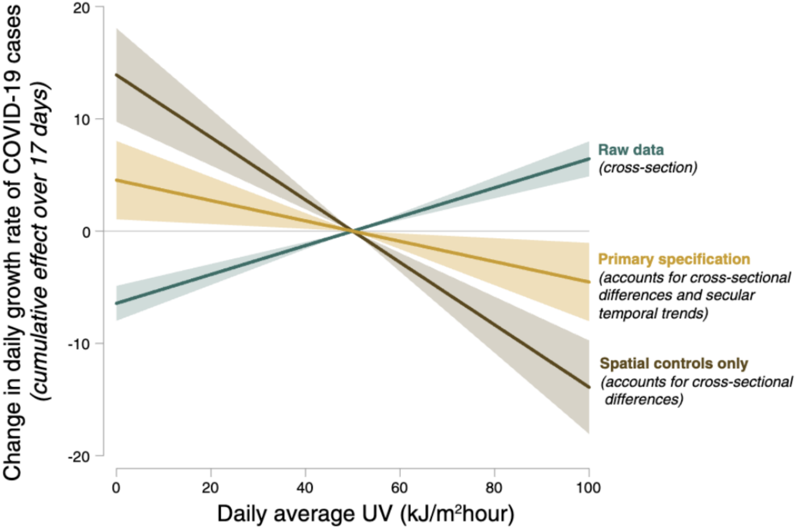
2. Effectively demean all observations within each location, i, so that f is estimated using only temporal variation within locations.

E.g., estimating the impact of climate on COVID transmission



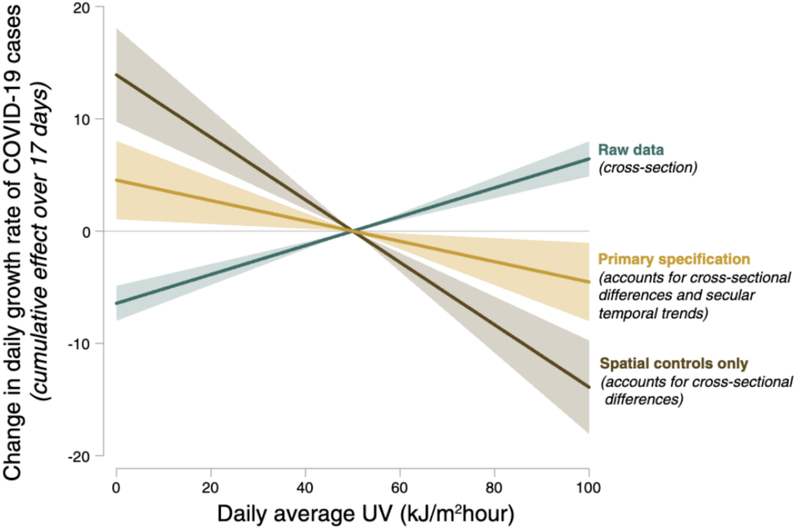
(Carleton et al., 2021)

E.g., estimating the impact of climate on COVID transmission



(Carleton et al., 2021)

E.g., estimating the impact of climate on COVID transmission



(Carleton et al., 2021)

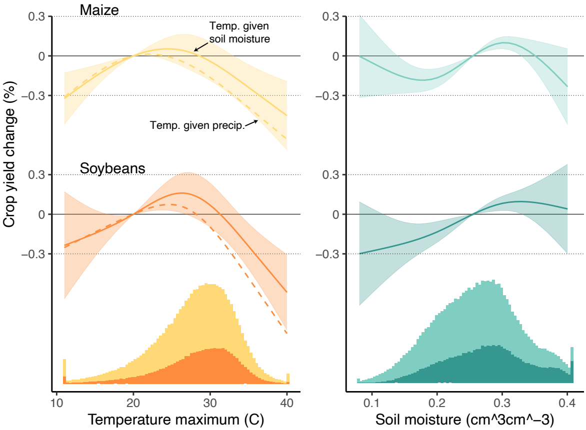
The spatial relationship is opposite the temporal relationship!

Summary: fixed effects isolate variation in the climate variable that is "as good as randomly assigned" so the effect of the climate variable on the outcome can be interpreted as causal.

Discussion

- These models commonly identify the effect of *weather*. Whereas what we are interested in is the effect of *climate*.
    - These could differ for a variety of reasons including human adaptation. People often estimate the effect of climate using lower-frequency variation in climate, though doing so can reduce the plausibility of the identifying assumption.
- Spatial fixed effects and trends cannot capture high-frequency time-varying variables that could confound the analysis.

The impact of temperature and soil moisture on crop yields



(Proctor et al., 2022)

“Projections using temperature and precipitation overestimate damages by 28% to 320% across crops because they confound stresses from dryness and heat.”

Estimating a local non-linear response using aggregated data

Challenge: we want to know the local, potentially non-linear, response of each individual unit (e.g. a single plant or human) so that we can make accurate projections of future impacts and their distribution.

But we often observe only aggregated (e.g., county, state, or national-level) observations of y.

Example: estimate the response of maize to temperature and soil moisture globally, with only national-level yield observations.

- If the local response were quadratic, and half of a nation's crops were below the optimal temperature and half were above the optimal temperature, then you might estimate no effect for small historical changes in temperature. But larger future increases in temperature could generate substantial net damages.
- Such a coarse response function would also miss the distributional impacts, with the cooler fields benefiting and the hotter fields losing from warming.

Estimating a local non-linear response using aggregated data

Idea: for outcomes that are aggregates (across space and/or time) the local non-linear response can be estimated by relating the aggregate outcome to aggregate non-linear temperature exposure, in a linear model.

Derivation

Consider observations of y (say maize production in tonnes/hectare) for a given county i and year t. We want to know the response of y at the the local level, say a 5km x 5km field, f.

$$y_{ft} = f(T_{ft})$$

And, let's assume that f takes a quadratic form:

$$y_{ft} = \beta_1 T_{ft} + \beta_2 T_{ft}^2$$

Now, sum over fields, f, within each county, i.

$$\sum_{f \in i} y_{ft} = \sum_{f \in i} \beta_1 T_{ft} + \sum_{f \in i} \beta_2 T_{ft}^2$$

Re-arraigning we get:

$$y_{it} = \beta_1 \sum_{f \in i} T_{ft} + \beta_2 \sum_{f \in i} T_{ft}^2$$

Derivation

$$y_{it} = \beta_1 \sum_{f \in i} T_{ft} + \beta_2 \sum_{f \in i} T_{ft}^2$$

Thus, the local, nonlinear f (β_1 and β_2) can be estimated by regressing the county-level production (y_{it}) on the aggregate county-level temperature exposure ($\sum_{f \in i} T_f$ and $\sum_{f \in i} T_f^2$), both of which are observable.

Derivation

$$y_{it} = \beta_1 \sum_{f \in i} T_{ft} + \beta_2 \sum_{f \in i} T_{ft}^2$$

Thus, the local, nonlinear f (β_1 and β_2) can be estimated by regressing the county-level production (y_{it}) on the aggregate county-level temperature exposure ($\sum_{f \in i} T_f$ and $\sum_{f \in i} T_f^2$), both of which are observable.

This approach works because of the linearity of the model

- non-linear transformation of T still allow f to be non-linear.

Derivation

$$y_{it} = \beta_1 \sum_{f \in i} T_{ft} + \beta_2 \sum_{f \in i} T_{ft}^2$$

Thus, the local, nonlinear f (β_1 and β_2) can be estimated by regressing the county-level production (y_{it}) on the aggregate county-level temperature exposure ($\sum_{f \in i} T_f$ and $\sum_{f \in i} T_f^2$), both of which are observable.

This approach works because of the linearity of the model

- non-linear transformation of T still allow f to be non-linear.

Generally, estimates are made at the resolution of the climate data, which tends to be finer resolution than the social outcome data.

Derivation

$$y_{it} = \beta_1 \sum_{f \in i} T_{ft} + \beta_2 \sum_{f \in i} T_{ft}^2$$

Thus, the local, nonlinear f (β_1 and β_2) can be estimated by regressing the county-level production (y_{it}) on the aggregate county-level temperature exposure ($\sum_{f \in i} T_f$ and $\sum_{f \in i} T_f^2$), both of which are observable.

This approach works because of the linearity of the model

- non-linear transformation of T still allow f to be non-linear.

Generally, estimates are made at the resolution of the climate data, which tends to be finer resolution than the social outcome data.

The approach can similarly be applied to temporally aggregated variables.

Practice question

True or False:

$$\sum_{f \in i} T_f^2 = (\sum_{f \in i} T_f)^2$$

I.e., is the sum of squares the same as the square of the sum?

Summary 1: fixed effects isolate variation in the climate variable that is "as good as randomly assigned" so the effect of the climate variable on the outcome can be interpreted as causal.

Summary 2: the local nonlinear response of the social outcome to the climate variable can be estimated by regressing the aggregate social outcome on aggregate non-linearly transformed climate variables.

Many models take a form similar to:

$$y_{it} = \beta_1 \sum_{f \in i} T_f + \beta_2 \sum_{f \in i} T_f^2 + \lambda_i + \phi_i t + \epsilon_{it}$$

Extensions

Heterogeneity:

- f likely differs across space, time, and other attributes.
- This can be modeled by including interaction terms in the regression. Below we allow the effect to differ by state, s:

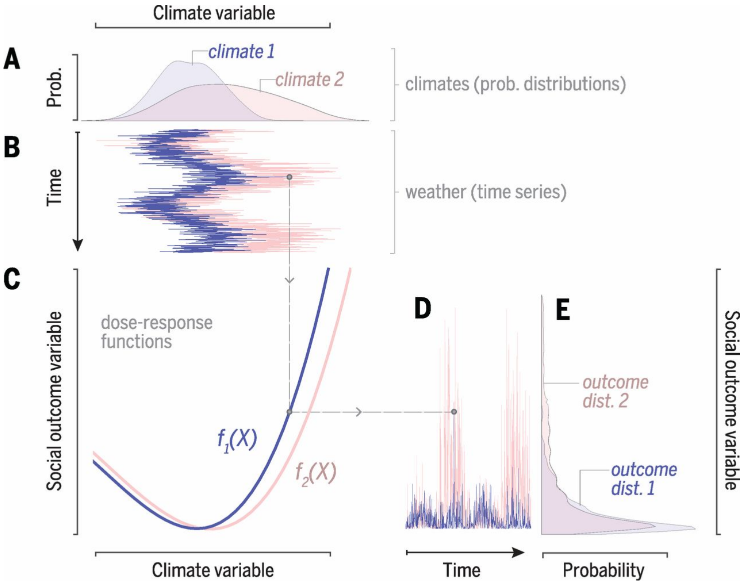
$$y_{it} = \beta_1 \sum_{f \in i} T_f + \beta_2 \sum_{f \in i} T_f^2 + \sum_s \mathbf{1}\{I \in S\}(\beta_1^s \sum_{f \in i} T_f + \beta_2^s \sum_{f \in i} T_f^2) + \lambda_i + ...$$

- High-complexity models can be combined with penalization (e.g., ridge) on the response function to avoid overfitting.

Model selection:

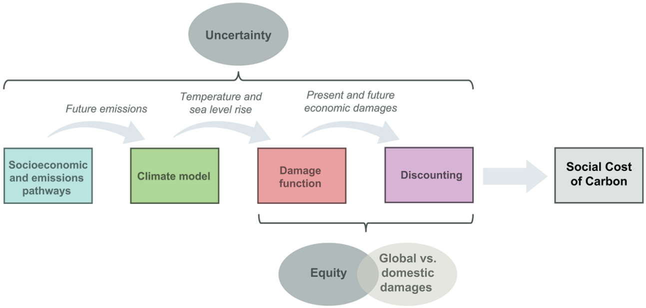
- Remove variation associated with the fixed effects prior to conducting model selection & evaluation.
- The fixed effects set the variation you want to analyze (i.e., the experimental design).
- Two step procedure is justified by the Frisch-Waugh-Lovell Theorem

Projecting climate impacts using dose-response functions



(Carleton & Hsiang, 2016)

Dose-response or "damage" functions provide the foundation for quantifying the "Social Cost of Carbon" (SCC)



(Carleton & Greenstone, 2022)

"damage estimates are not currently available for all sectors affected by climate change...many difficult-to-quantify sectors, such as ecosystem services and human migration, are not included"

Lots of opportunity for novel, policy-relevant research

- Understanding how the climate influences the many facets of human wellbeing is critical for anticipating, adapting to, and mitigating future climate risks; but there's still so much we don't know.
- Growing availability of data (e.g., satellite observations, crowd-sourcing, censuses), increasing computational resources, and advances in function approximation make this a field ripe for investigation.
- If you're interested in exploring these topics further, I'm currently accepting applications for graduate students to my lab at UBC. See `https://www.jonathanproctor.org/join` for details.

Open-ended question

- What is an outcome (e.g., agricultural productivity, sleep quality, human migration, labor productivity...etc) that you think might be affected by climate change?
- How would you estimate that effect?