Northwestern | INSTITUTE FOR POLICY RESEARCH

RESEARCH EXCELLENCE • POLICY IMPACT

IPR Working Paper Series

WP-26-14

With or Without U? Binning Bias and the Causal Effects of Temperature Extremes

Benjamin Jones
Northwestern University and IPR

Jacob Moscona
Massachusetts Institute of Technology

Benjamin Olken
Massachusetts Institute of Technology

Cristine von Dessauer
Massachusetts Institute of Technology

Version: June 11, 2026

DRAFT

Please do not quote or distribute without permission.

Abstract

Estimates of climate impacts show that extreme temperatures have large and widespread effects. To estimate these effects, a common approach counts days in different temperature ranges and considers how exposure to these distinct "bins" affects outcomes. This often produces non-linear, U-shaped results, in which high and low temperatures have the largest effects. The authors show that nonlinear approaches like these can generate spurious findings. Specifically, global warming induces trends in extreme temperature exposure that correlate mechanically with a location's baseline temperature. Substantial bias emerges if trends in the outcome variable also correlate with baseline temperature for any reason. The authors demonstrate this problem theoretically, in simulations, and with real outcomes. They then develop solutions. In applications using US data, some results in the literature are unaffected by these corrections, while other results change substantially.

Acknowledgements

The authors thank Isaiah Andrews, Olivier Deschenes, Raffaele Ferrari, Peter Hull, Peter Huybers, Jesse Shapiro, Joe Shapiro, Jonathan Roth, Wolfram Schlenker, Andy Solow, Jim Stock, and Mengze Wang for helpful comments and suggestions. They thank Harufumi Nakazawa for outstanding research assistance. This project was supported in part by the MIT Climate Project. Benjamin A. Olken is a director of J-PAL at MIT, the Co-Scientific Director of J-PAL Southeast Asia, and CoDirector of J-PAL's Social Protection Initiative. He serves as the Co-Director of the Development Economics Program for the National Bureau of Economic Research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

1 Introduction

Amid rising attention to climate change and concern about the potantial economic impacts of global warming, researchers have estimated the impact of temperature on a wide range of economic outcomes. Since temperature variation across places can be correlated with other factors, economists have turned to panel models to identify temperature's effects (e.g., Deschênes and Greenstone 2007; Dell et al. 2012; see Dell et al. 2014 for a review). The panel approach leverages year-to-year or decade-to-decade fluctuations in temperature, within a given place, to isolate idiosyncratic temperature variation and reveal its impact on outcomes of interest.

When implementing panel approaches, researchers often consider potential nonlinearities, where exposure to temperature extremes could be especially important. One common approach, for example, is to count the number of days a location experiences each year in distinct temperature bins (e.g., the number of days $> 90°$F, the number of days $80-90°$F, and so on) and estimate a panel regression with these day counts as regressors. A recurrent finding across a broad range of outcomes is that temperature damages are driven by extreme hot days and/or cold days, often generating "U-shaped" or "inverse-U-shaped" effects (e.g., Ranson, 2014; Graff Zivin and Neidell, 2014; Barreca et al., 2016; Hsiang et al., 2017; Park et al., 2020; Cohen and Dechezleprêtre, 2022; Carleton et al., 2022). For example, exposure to both very hot and very cold days have been shown to substantially lower crop yields (e.g., Schlenker and Roberts, 2009) and increase mortality (e.g., Deschênes and Greenstone, 2011). These non-linear estimates have become influential; for example, they have been used to assess the social cost of carbon (EPA, 2023).

In this paper, we consider a serious bias that can emerge in these regressions. Specifically, in the presence of general warming, extreme hot and cold temperatures will, by construction, exhibit place-specific trends that are strongly correlated with each location's initial temperature. If the outcome variable also has place-specific trends that are correlated with the location's initial temperature for any reason, the extreme temperature coefficients become biased. That is, the panel model is no longer identified by idiosyncratic temperature shocks; instead, for extreme temperatures, one is in effect regressing trends on trends. While we focus on binned temperature regressions in our main analysis for analytical simplicity, the bias we discuss here is also an issue for other non-linear transformations of daily temperature data, including polynomials of daily temperatures summed across time and so-called 'killing degree days' (e.g., Butler and Huybers 2013).

To see the issue intuitively, suppose that climate change is uniform, so that the mean temperature in each location increases by the same amount each year. Different locations, however, have different initial temperatures. This implies that the impact of overall warming on exposure to each part of the temperature distribution will differ across space. For example, a hot place like Phoenix, AZ, will experience a large increase in the number of very hot days (e.g., days with a mean temperature above 90°F), whereas a cold place, like Boston, MA, will experience little such increase. These very hot days occur only in the upper tail of Boston's temperature distribution but are in a thicker part of Phoenix's distribution, as illustrated in Figure 1. The implication is that Phoenix, relative to Boston, will have a steeper upward trend in its count of very hot days.

The same argument holds for very cold days. With general warming, a cold place like Boston will experience a decline in the number of very cold days – say, days with a mean temperature below 10°F – whereas Phoenix never had such days to begin with. The result is that Phoenix, compared to Boston, will exhibit a differential upward trend in the number of very cold days (i.e., Phoenix has a less steep negative trend, which is equivalent to a differential upward trend). In sum, Phoenix will appear to have a differential upward trend in both very hot days and very cold days. Meanwhile, there will be little difference in trends in intermediate temperatures, which are in the thick part of the temperature distribution for both cities.

In this way, generalized warming creates differential trends in the hottest and coldest temperatures, driven by how rising temperatures interact with differences in baseline temperature. If the outcome variable also has differential trends that are correlated with baseline temperature – for example, if hot parts of the country evolve in different ways from colder parts of the country, as may be the case for completely unrelated reasons – this generates bias. For example, let's say that with the advent of air conditioning, population has been moving south in the U.S., to places that were initially hot. Then, taking the Phoenix and Boston logic above, an initially hot location's differential upward trend in population will be positively correlated with the hot location's differential upward trends in hot and cold bins, producing a U-shaped pattern in the panel regression. The econometrician would then conclude that a year with more very hot days attracts inward migration, while a year with more very cold days also attracts more inward migration. This bias, driven by correlated differential trends, can thus generate U-shaped (or inverse-U-shaped) coefficient patterns – shapes like those observed in the literature. This occurs even if yearly temperatures have no causal effect on outcomes whatsoever.

This paper explores this potential bias – which we term *binning bias* – in three steps. In the first step, we document the econometric problem, using both theory and a series of simulations. The theoretical treatment examines simple data generating processes and develops explicit expressions for binning bias, demonstrating conditions and logic for biased coefficients, including the "U-shaped" patterns that are common in existing work. The simulations then demonstrate binning bias empirically, starting with synthetic data and moving to real world temperature and outcome data. While we work primarily with discrete bins of temperature, we also use simulations to show that the same bias arises in other common nonlinear transformations of temperature.

The initial simulations consider synthetic outcomes and temperature processes. Temperatures are drawn from normal distributions that differ in their means across locations. Outcomes are *iid* random variables with place-specific trends but no relationship with contemporaneous temperatures. When there is uniform warming and outcome trends are correlated with a location's baseline temperature, binned panel regressions generate spurious U-shaped patterns as described above. That is, we would falsely conclude that extreme hot and cold days in a given year had an impact on the outcome of interest, even though – by construction – there is no such effect. We can generate either U-shapes or inverted U-shapes, consistent with the theory, simply by switching the correlation of the outcome trends with initial temperatures.

The next set of simulations turn to real-world temperature data, which has more complicated features and dynamics than the simulated temperature distributions in the first simulation. We use U.S. county-level temperature data since 1980. Our main simulations use mean daily temperatures constructed from ERA-5-Land (Muñoz-Sabater et al., 2021), though we also conduct simulations using other temperature series such as harmonized PRISM data (Schlenker, 2024) and station-level data from GHCN (Menne et al., 2012), with similar results. When we continue to use synthetic outcome variables with trends that correlate with each county's baseline temperature, binned panel regressions again find spurious U-shaped effects. By contrast, when the outcome variable trends are uncorrelated with baseline temperatures, the bias disappears. This finding underscores that a correlation between outcome trends and baseline temperature is necessary to generate spurious effects, emphasizing that the relevance of binning bias is contingent on the dynamics in the dependent variable and will thus depend on the outcome of interest.

To gauge the potential magnitude of bias in common applications, we further calibrate outcome trends to real-world data, drawing on outcomes that have been studied in prior

work: mortality rates, crime rates, and crop yields. The goal is to investigate whether time trends in real data are sufficiently large in magnitude, and sufficiently correlated with baseline temperatures, to generate levels of binning bias that are quantitatively meaningful. We find that in all cases real-world time trends are sufficient to generate binning bias, while the issue is more severe in some cases (e.g., mortality) than others (e.g., crop yields). While these estimates *do not* yet imply that existing estimates are biased, they *do* show that the bias we are investigating is quantitatively large enough to potentially generate effects similar to the size reported in the literature.

We further study "adaptation." In particular, many studies argue that places with pre-existing exposure to extreme heat or extreme cold are likely to be more adapted to these conditions. Thus, smaller effects of extreme heat in hot locations (or of extreme cold in cold locations) are interpreted as evidence of adaptation and used to inform future climate damages (e.g., Carleton et al., 2022; Hultgren et al., 2022). However, binning bias may undermine this inference. Repeating our simulations separately for cold and hot counties, we find larger spurious effects of hot days in colder counties and cold days in hotter counties. Further, this finding is what one would expect when applying the bias formulas from the theoretical treatment. Empirical patterns often interpreted as "adaptation" may then be an artifact of binning bias.[1]

In light of these challenges, the second part of our paper explores potential solutions. One approach focuses on the problematic dynamics in the binned temperature variables. For causal identification, we need the identifying variation to be based on idiosyncratic shocks. This suggests that if we can model the underlying temperature data generating process, we can control directly for the expected number of days in each bin, place, and time. Such expected bin counts provide counterfactual controls, accounting for any trends induced by the binning process and allowing the estimates to identify causal effects from idiosyncratic variation alone. This counterfactual-expectation control approach is in the same spirit as suggestions by Wooldridge (2015) and Borusyak and Hull (2023).[2]

Implementing this procedure, however, requires constructing the counterfactual set of possible temperature realizations that could have been realized in a given place and time. This is not trivial, since each location may have different distributions of temperatures, and constructing place-and-time specific distributions of potential temperatures is an

[1]Heterogeneity analyses may more generally exhibit bias if the margin of heterogeneity is correlated with differences in trends or their variance.

[2]See in particular Borusyak and Hull (2023) section 3.2, which suggests controlling for the expectation of each variable based on simulations.

active area of research in climate science (see, e.g., Wang et al., 2024). We proceed with a empirical approach that uses the realizations of all temperatures over the entire period to construct counterfactuals for each location with non-parametric distributions.[3]

Our procedure for constructing counterfactuals is as follows. We take the realized distribution of daily temperatures separately for each location-month.[4] For each location-month, we estimate a linear trend in the mean monthly temperature over the years studied, which we shrink using standard empirical Bayes techniques. We then recenter each location-month distribution by subtracting off its estimated mean shift in any given year and pool these re-centered distributions. With 40 years of data for each location and 30 days in a month, for example, the location-month distribution consists of 1,200 re-centered data points. Finally, we project the location-month's entire temperature distribution forward each year, shifting the distribution uniformly to the right according to the location-month-specific time trend. Once we have this predicted distribution in each year, we count the expected number of days in each temperature bin in each location-year and add these expected counts as control variables to the panel regression.[5]

We use our simulations—with observed temperature variables but simulated outcome variables—to test whether this counterfactual control approach eliminates the bias. We find that the counterfactual approach is successful. Moreover, when we add causal effects of extreme heat or cold to the data generating process, the counterfactual approach successfully purges the binning bias and correctly estimates the true causal effects. We further consider and present numerous alternative counterfactual approaches to modeling local temperature distributions. The relatively straightforward approach outlined above works especially effectively in U.S. data when compared to other approaches.

We also evaluate several reduced-form methods. First, we consider other approaches common in the literature, including controlling for region-specific time dummies (e.g., state-by-year fixed effects), additional lags of left and/or right-hand side variables, and place-specific linear trends. Second, motivated by the theoretical result that binning bias is more severe in long panels, we investigate the impact of including county fixed effects interacted with short (e.g., five year) calender year periods, to effectively use repeated short panels for identification. This 'short panel' approach isolates variation across years within narrower time windows where idiosyncratic variation is more likely to dominate

[3]We thank Peter Huybers for suggesting this data-driven approach to constructing counterfactuals.

[4]We work separately by month to allow flexibility for seasonality, which may differ by location.

[5]We provide code to create these counterfactuals using the *cftemp* Stata package, which is available on our websites.

variation driven by temperature trends. We evaluate these approaches using our baseline simulated outcome with place-specific linear trends and with alternative place-specific outcome dynamics (e.g., quadratics).[6]

We find that state-by-year fixed effects and lagged-variable approaches do not eliminate the bias. Controlling for place-specific linear trends or five-year bin indicators interacted with county fixed effects (i.e. the "short panel" approach), however, work in some contexts and can be useful for gauging the extent of binning bias in applications, but are not always applicable.[7] Overall, the counterfactual approach, which works to solve the problem on the right-hand side of the regression, appears to have quite general applicability, though the 'short panel' approach is a useful reduced-form alternative when applicable.

The final step of our paper revisits a number of results in the literature. We examine four sets of outcome variables: county-level population counts in each decade from the U.S. Census, annual mortality rates at the county-level from the Centers for Disease Control (CDC), county-level crop yields from the U.S. Department of Agriculture National Agricultural Statistics Service (NASS), and county-level crime data from the FBI compiled by Ranson (2014). Population movement, mortality, agricultural productivity, and crime have all been major areas of focus in existing work estimating the economic impacts of temperature change, and each has been proposed as a major contributor or adaptation mechanism to climate damages. For each outcome variable, we estimate the impact of temperature using the binned specification with county and year fixed effects and then consider what happens to these estimates when we correct for binning bias.

We find that binning bias has the potential to affect substantive conclusions, but results vary depending on the application. For population counts, our naive estimates generate a U-shaped relationship whereby both extreme heat and extreme cold are associated with population increases. After correcting for binning bias, however, the magnitudes fall dramatically, and we observe (if anything) a negative effect of exposure to the top temperature bin. This could be evidence of population movement away from extreme heat; however, naive estimates would instead produce the opposite result. For mortality,

[6]Since in our benchmark simulations we generate a left-hand side variable with place-specific linear trends, clearly including place-specific linear trends will solve the problem, but there is no reason to suppose that the place-specific trends are in fact linear in the real world; we therefore consider quadratic trends, and other functional forms to check the robustness of these various solutions.

[7]Specifically, place-specific linear trends work if the trend in the outcome variable is linear but will not work as well when the trends are non-linear. The short panel approach appears more robust to different functional forms of the trends, but cannot be used in settings where the outcome is not annual (i.e. with decadal data) or when researchers want to estimate effects over longer time horizons.

we find that the size of the U-shaped relationship between temperature and mortality falls substantially when we account for binning bias. Extreme heat no longer raises all-cause mortality and extreme cold has a positive but smaller effect. It is not always the case, however, that correcting for binning bias dampens or reverses existing estimates. For crop yields, we find the negative effects of high temperatures is unaffected by binning bias. And for violent crime, we find that accounting for binning bias accentuates the negative relationship between cold days and violent crime. This implies a larger potential effect of warming on violent crime than would be uncovered absent the bias correction.

The point of these exercises is not to establish the final word on any of these outcomes. Rather, it is to illustrate that binning bias can meaningfully affect empirical estimates by masking real effects or generating spurious effects. More generally, we suggest that accounting for this bias is an important issue for authors to consider when estimating the effect of temperature on economic outcomes.

The remainder of the paper is organized as follows. Section 2 presents a simple theory for binning bias, elucidating its logic and key forces that govern the bias. Section 3 describes the data we use and illustrates how place-specific trends in temperature bin variables will emerge mechanically when there is general warming. Section 4 presents simulations, showing how binning bias can generate the U-shaped patterns frequently observed in this literature. Section 5 introduces and evaluates potential solutions. Section 6 examines applications of our solution to population counts, mortality, crop yields, and crime. Section 7 concludes.

2 Theoretical Framework

In this section, we provide a formal analysis of binning bias. We develop explicit conditions on when this bias will occur and analyze cases where U-shapes, inverted U-shapes, or other forms of non-linear bias will emerge.

To begin, consider a standard panel regression featuring temperature bins. The idea is that hot days, temperate days, and cold days may have quite different effects on outcome variables of interest. This possibility is often analyzed in models of the following form:

$$Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k \times T_{kit} + u_{it} \tag{1}$$

where Y_{it} is the outcome variable for place i at time t, α_i are place-specific fixed effects, λ_t are time fixed effects, and T_{kit} represent counts of days in temperature bin k in a given time and place. The bins consider different ranges of temperature, often 10°F or 5°C wide. The β_k are intended to provide the treatment effects of interest, estimating how more days in any given bin affect the outcome.

2.1 The Outcome Variable

First consider the data generating process for the outcome variable. We will consider the case where researchers run a panel model as in (1) but the true data generating process for the outcome variable is

$$Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k \times T_{kit} + \gamma_i t + \varepsilon_{it} \tag{2}$$

where the error is *iid* and $\mathbb{E}[\varepsilon|t, T_{kit}] = 0$. The key here is the introduction of the $\gamma_i t$, which bring in potential place-specific trends in the outcome variable (and are omitted in the regression specification (1)). These trends may covary with other place-specific characteristics and are independent of the error terms, $\mathbb{E}[\gamma_i \epsilon_{it}] = 0$.

2.2 The Temperature Bins

Now consider the data generation process for the temperature bins. For simplicity, consider $k = 3$ temperature bins, where C_{it} is a count of cold days, M_{it} is a count of moderate temperature days, and H_{it} is a count of hot days. Let periods be one year. The sum of the day counts thus add up to 365 days in any given place and year; given this collinearity, we drop the count of moderate days, M_{it}, from the regression. Estimates of treatment effects of C_{it} and H_{it} will be interpreted as differences from the effect of moderate days.

Let the true data generating process for counts of cold days and hot days be

$$C_{it} = \mu_{C,i} + \delta_{C,i} t + v_{C,it} \tag{3}$$

$$H_{it} = \mu_{H,i} + \delta_{H,i} t + v_{H,it} \tag{4}$$

where the error terms are *iid* with $\mathbb{E}[v_{.,it}|t] = 0$, $\mathbb{E}[\delta_{.,i} v_{.,it}] = 0$, and variances σ_C^2 and σ_H^2 respectively.[8]

[8]Note that, since total days in each year add up to 365, the bin-level error terms cannot be fully

As with the dependent variable (see equation (2)), we have introduced the potential for place-specific differential trends into these temperature variables, represented here by $\delta_{C,i}t$ and $\delta_{H,i}t$. As described in the introduction, these differential trends are a natural consequence of global warming. In particular, if we think of warming as shifting the entire distribution of temperature in a given place to the right, we expect fewer cold days with time ($\delta_{C,i} < 0$). But since initially hot places experience few if any cold days to start with, this decline will inevitably be much less negative in initially hot places. Similarly, with warming, we expect more hot days with time ($\delta_{H,i} > 0$) but this increase will inevitably be much smaller in initially cold places, which may still see very few very hot days (see Figure 1). Thus these day counts will tend to have differential trends depending on the baseline temperature of the place.

2.3 Forms of Bias

We can now consider biases that can result when econometricians implement a traditional panel model. Namely, consider the panel regression

$$Y_{it} = \alpha_i + \lambda_t + \beta_C C_{it} + \beta_H H_{it} + u_{it} \tag{5}$$

but where the true data generating processes are given by (2), (3), and (4).

Given the outcome variable's data generating process, the error term in (5) is

$$u_{it} = \gamma_i t + \epsilon_{it}. \tag{6}$$

Bias can then result when estimating (5) should the trending terms in either C_{it} or H_{it} have any covariance with these error trends. Intuitively, time fixed effects cannot solve this problem because, while they account for common dynamics across all places, they do not account for these differential trends by place. The resulting biases are derived as follows.

Result 1. *For large T, the biases in the coefficient estimates are*

$$Bias(\hat{\beta}_C) = \frac{Var(\delta_{H,i})\,Cov(\delta_{C,i}, \gamma_i) - Cov(\delta_{C,i}, \delta_{H,i})\,Cov(\delta_{H,i}, \gamma_i)}{Var(\delta_{C,i})\,Var(\delta_{H,i}) - \left(Cov(\delta_{C,i}, \delta_{H,i})\right)^2} \tag{7}$$

independent across all bins. Here we let the errors for cold and hot days be *iid*, with the moderate days being determined from these other independent processes. Extensions to the baseline model might consider non-independent errors.

$$Bias(\hat{\beta}_H) = \frac{Var(\delta_{C,i})\,Cov(\delta_{H,i},\gamma_i) - Cov(\delta_{C,i},\delta_{H,i})\,Cov(\delta_{C,i},\gamma_i)}{Var(\delta_{C,i})\,Var(\delta_{H,i}) - \left(Cov(\delta_{C,i},\delta_{H,i})\right)^2}. \tag{8}$$

These biases admit U-shape patterns (both biases are positive), inverted U-shape patterns (both biases are negative), or non-U-shape patterns, according to conditions summarized in Figure 2.

Proof See Appendix.

Note that, from this result, the existence of place-specific trends in the outcome and temperature bin variables does not necessarily imply bias. In particular, there will be no bias if the place-specific trends in the outcome variable are not covariant with the place-specific trends in the temperature variables (i.e., if $\text{Cov}(\delta_{C,i},\gamma_i) = 0$ and $\text{Cov}(\delta_{H,i},\gamma_i) = 0$). However, a sufficient condition for bias is simply that at least one of the temperature bin variables shows differential trends across place that are correlated with differential outcome trends.

The bias formulas in equations (7) and (8), and the result in Figure 2, allow for arbitrary correlations between trends in the temperature bins and trends in the outcomes. But as discussed above, we are motivated by a particular case where place-specific trends are correlated with initial temperature—i.e., where hot and cold day counts will have trends that are correlated by construction with a place's initial temperature (see discussion above). In this case, we can derive more specific results.

Specifically, suppose there is a relationship between trends in the temperature bins and initial temperature of a given place i, which we denote $T_{0,i}$. Write:

$$\delta_{C,i} = \psi_C + \omega_C T_{0,i} + e_{C,i} \tag{9}$$

$$\delta_{H,i} = \psi_H + \omega_H T_{0,i} + e_{H,i} \tag{10}$$

$$\gamma_i = \psi_Y + \omega_Y T_{0,i} + e_{Y,i} \tag{11}$$

where the error terms $e_{.,i}$ are all *iid* with variances $\sigma^2_{e_.}$. Let the variance of $T_{0,i}$ be $\sigma^2_{T_0}$. With warming, we expect $\psi_C < 0$, so that cold days are on average trending downwards, and $\omega_C > 0$, so that places with initially low temperature will have the most rapid decline in the number of cold days. Similarly, with warming, we expect that $\psi_H > 0$ so that hot days are on average trending upwards, and $\omega_H > 0$ so that places with initially low temperature will have smaller increases in the number of hot days. We then have the following result:

Result 2. *With the trends defined in (9), (10), and (11), the biases in the coefficient estimates are*

$$Bias(\hat{\beta}_C) = \frac{\omega_C \omega_Y}{\omega_C^2 + \frac{A_C}{A_H}\omega_H^2 + A_H} \tag{12}$$

$$Bias(\hat{\beta}_H) = \frac{\omega_H \omega_Y}{\omega_H^2 + \frac{A_H}{A_C}\omega_H^2 + A_C} \tag{13}$$

where $A_C = \frac{\sigma_{e_C}^2}{\sigma_{T_0}^2} + \frac{T-1}{S_t}\frac{\sigma_C^2}{\sigma_{T_0}^2}$, $A_H = \frac{\sigma_{e_H}^2}{\sigma_{T_0}^2} + \frac{T-1}{S_t}\frac{\sigma_H^2}{\sigma_{T_0}^2}$ *and* $S_t = \sum_{t=1}^{T}(t-\bar{t})^2$. *The bias is U-shaped if* $\omega_Y > 0$, *the bias is inverted U-shaped if* $\omega_Y < 0$, *and there is no bias if* $\omega_Y = 0$.

Proof See Appendix.

Result 2 elucidates several features of binning bias that we will explore in the next sections. As a general intuition, the bias problem will naturally be worse when differential trends on both sides of the regression are more tightly related. These trend relationships depend on the trends' correlations with baseline temperature (the ω parameters) and the degree of noise (the σ^2 parameters).

Consider four key features, focusing on the case where $\omega_Y > 0$ (generating a U-shape). First, bias increases in ω_Y. Intuitively, the larger the correlation between place-specific outcome trends and initial temperature, the greater the bias will be. Second, the bias in $\hat{\beta}_C$ declines as $\sigma_{e_C}^2$ rises. Intuitively, as a temperature bin's trends become driven more by random noise than a systematic relationship to initial temperature, the bias for the bin is attenuated. Third, and perhaps counter-intuitively, longer panels tend to make binning bias *worse*. The reason is that idiosyncratic inter-annual variation (via σ_C^2 and σ_H^2), which provides the shocks to each bin's day counts and which are intended to identify temperature effects, becomes dominated by any trend the longer those trends proceed.[9] Thus longer panels tend to amplify the binning bias problem.

Finally, the effect of ω_C on the bias in $\hat{\beta}_C$ is more subtle. On the one hand, a smaller slope with initial temperature (smaller ω_C) causes the signal to noise ratio to go down, which works to reduce the bias (the same way that increasing the noise, $\sigma_{e_C}^2$, reduces bias). On the other hand, a smaller slope with initial temperature (smaller ω_C) means that smaller variations in day counts can now predict the given variation in the outcome, resulting in larger bias. The latter force will actually dominate when noise is relatively low (e.g., in longer panels or with small $\sigma_{e_C}^2$). Overall, small but precise relationships

[9] In Result 2, as T increases, the A_C and A_H terms decline, since T/S_T is decreasing in T.

between temperature bin trends and initial temperature are especially pernicious, driving large bias. We will see these features in practice in the empirical analyses to come.

3 Data and Differential Trends in Temperature Bins

In this section, we first describe our main data sources and describe the binning procedure. We then provide concrete demonstrations of how binning the temperature data can lead to differential trends in these variables, especially for temperature extremes.

3.1 Data Sources

Temperature Our main analyses use temperature data from the ERA5-Land database (Muñoz-Sabater et al., 2021). This reanalysis data set combines weather observations from around the world with a model to generate gridded (0.1-by-0.1 degrees), hour-by-hour measurements since 1970. We also replicate our main results using alternative sources of temperature data, including daily U.S. county-level data from Schlenker (2024), which uses a balanced panel of weather stations from the PRISM database, as well as raw station level data from the Global Historical Climatology Network (GHCN) (Menne et al., 2012).

We use the hourly temperature data to construct the average temperature in each grid cell and day, focusing only on daytime hours (7am-6pm). We then assign each U.S. county the grid cell that is geographically nearest to its centroid and, for each county and year, count the number of days that fall within distinct temperature bins: below 10°F, 10°-20°F, 20°-30°F, 30°-40°F, 40°-50°F, 50°-60°F, 60°-70°F, 70°-80°F, 80°-90°F, and over 90°F. We use the number of days that fall within each bin in each county-year as our main measure of exposure to moderate and extreme temperatures.

Outcome Data We further collect outcome data for the analyses of specific applications in Section 6. In brief, to measure population changes across U.S. counties, we compile data from each round of the U.S. Census from 1970 to 2010. To measure mortality rates by county and age group, we compile data from the Centers for Disease Control (CDC) compressed mortality files from 1970-2002. This database contains records of the universe of U.S. deaths, broken down by year, county of residence, and age group (among other characteristics).[10] To measure crop yields, we use data from the U.S. Department of

[10]We use the public version of these data, which omit death counts in specific age ranges in specific county-years when a particular category of deaths are rare. We focus on over-65 mortality, for which

Agriculture National Agricultural Statistics Service on yearly output in each U.S. county of corn, wheat, and soybeans from 1985 to 2006. To measure crime, we use data compiled from FBI reports by Ranson (2014), covering years from 1960-2009. We count the number of violent crimes—including murder, rape, and aggravated assault—and non-violent or property crimes—including robbery and larceny—in each U.S. county, month, and year.

3.2 Illustration of differential trends in temperature bins

Using the county-level temperature dataset, we first illustrate how binning the temperature data introduces strong relationships between exposure to temperature extremes and baseline temperature. Consider, for example, two U.S. cities: Boston and Phoenix. Table 1 shows counts of hot and cold days in each place during the 1970s and 2010s. During the 1970s, Boston had 20 days with an average temperature below 10°F. In the 2010s, this number had dropped to 10. Meanwhile, in both decades, Boston experienced zero days with a mean daytime temperature above 90°F. Thus, global warming led Boston—a cold place—to experience a large decline in exposure to the bottom temperature bin and no change in exposure to the top temperature bin. Phoenix, by contrast, is a hot place. In both the 1970s and 2010s, Phoenix had zero days below 10°F, while the number of days over 90°F increased from 1,000 in the 1970s to 1062 during the 2010s. Phoenix thus experienced no change in exposure to the bottom temperature bin yet a large increase in exposure to the top temperature bin. The different baseline temperatures of the two cities meant that—as temperatures increased—their changes in exposure to both ends of the temperature distribution would be uneven.

This pattern is true more systematically and generates a mechanical, positive relationship between county-level baseline temperature and measured changes in exposure to extreme temperatures across all U.S. counties. Figure 3 plots the cross-county relationship between average temperature during the 1970s (x-axis) and changes in exposure to cold (below 10°F) and hot (above 90°F) days (y-axis). Both trends are strongly upward sloping ($p \leq 0.001$). On average, colder counties mechanically experience larger declines in exposure to extreme cold, leading to a positive correlation between baseline temperature and the change in exposure to extreme cold; meanwhile, hotter counties mechanically experience larger increases in exposure to extreme heat, leading to a positive correlation between baseline temperature and the change in exposure to extreme heat. Exposure

there is only 0.6% of the observations missing.

to moderate temperature, on the other hand, does not strongly correlate with baseline temperatures, since all counties had substantial exposure to moderate temperatures at baseline, meaning that warming had a comparable effect on their changes in exposure.

Extreme temperature bins—commonly used variables in empirical analysis that are often assumed to be exogenous—thus mechanically have trends that are strongly correlated with baseline temperature. The next section will show, consistent with the theory, that such differential trends can generate spurious regression results.

4 Simulations and Binning Bias

In this section, we use a series of simulations to study how non-linear transformations of temperature can lead to spurious effects of temperature extremes on outcome variables. The main finding is that binning daily temperature realizations or using other non-linear functional forms can lead to "U" or "inverse-U" shaped results, even when there is no causal relationship between the outcome variable and contemporaneous temperature, so long as the trend in the outcome is correlated with baseline temperature.

4.1 Simulated data with uniform temperature change

Our first simulations consider synthetic temperature data and synthetic outcome data. We generate temperature data with uniform warming, allowing for a 1.8°F (i.e., 1°C) increase in the temperature of all U.S. counties over a 50-year period.[11] Specifically, we take the actual mean temperature of each county c in 1970, $Temp_c^{Base}$, and let these means rise with time as follows:

$$\widetilde{\mu}_{ct} = Temp_c^{Base} + \frac{1.8}{50} \cdot (t-1) \tag{14}$$

where $t \in \{1, 50\}$. For each county and year, we draw 365 daily temperatures from a normal distribution centered at $\widetilde{\mu}_{ct}$ with a common fixed variance. We then count the number of days in each county-year in each temperature bin, defining the bins in 10°F intervals (less than 10°F, 10-20°F, etc.). This follows standard practice in the empirical literature and further introduces many temperature bins, allowing us to see how the degree of bias varies across more or less extreme temperature categories.

[11] This 1.8°F increase is roughly the change in average U.S. temperature over the sample period.

The outcome variable is simulated to have a linear trend correlated with baseline temperature but no relationship with contemporaneous temperature bins:

$$Y_{ct} = t \cdot Temp_c^{Base} + \epsilon_{ct} \tag{15}$$

where ϵ_{ct} is drawn from a normal distribution centered at zero and with variance σ_ϵ^2.[12]

With these data, we implement the standard panel model above, which is repeated here for convenience:

$$Y_{ct} = \alpha_c + \lambda_t + \sum_k \beta_k \times T_{kct} + u_{ct} \tag{1}$$

where c indexes counties, t indexes years, k indexes temperature bins (less than 10°F, 10-20°F, etc.), T_{kct} are the counts of days in each bin in each county and year, and α_c and λ_t are place and time fixed effects.

For 1,000 draws of the simulated outcome and temperature realizations, we estimate model (1). The estimates of β_k are presented in Figure 4a, where each dot represents the median coefficient estimate and the bars contain 95% of the coefficient estimates across all draws. Strikingly, the coefficient estimates in Figure 4a imply large positive effects of exposure to both hot and cold days despite there being *no* relationship between contemporaneous temperature extremes and the outcome. A uniform, rightward shift in the temperature distribution—combined with place-specific trends in the outcome that are correlated with baseline temperature—generates a spurious U-shaped relationship.

Intuitively, when the trend in the outcome is *negative* (i.e., t is replaced with $-t$ in equation (15)) the panel model generates a spurious inverted-U-shaped relationship between temperature exposure and the outcome (Figure 4b). That is, binning bias can generate either U-shapes or inverted U-shapes, depending on the correlation of outcome trends with baseline temperature. Finally, consistent with the theory, when there are no differential trends in the outcome the spurious effects are absent (Figure 4c).

This first exercise—with both simulated temperature and simulated outcomes—suggests that non-linear transformations of the temperature variable can generate spurious effects of extreme heat and extreme cold. Even with a uniform trend in temperature across locations, binning generates differential trends in the independent variables that are correlated with baseline temperature and are not absorbed by the time fixed effects. In the

[12] We set σ_ϵ^2 to be equal to the variance of $t \cdot Temp_c^{Base}$ over the entire sample, so that the error term in the outcome is of similar variance to trends, but this is not essential.

next section, we show similar patterns using real-world temperature data.

4.2 Real temperature data

Our next set of simulations use real temperature data. Observed temperature realizations present more complex dynamics than the synthetic data analyzed above with uniform warming. In this section, we investigate whether real temperature dynamics lead to similar spurious effects of temperature extremes. To do so, we construct the binned temperature variables from U.S. counties' actual temperature data while again considering synthetic outcome data – where, by construction, there is no contemporaneous effect of temperature on outcomes.

Specifically, using the ERA-5-Land temperature data, we count the days in each county-year that fell into each temperature bin (see Section 3.1). Thus, we now measure the *actual* exposure of each county in each year to all parts of the temperature distribution. We then simulate the outcome with linear trends in baseline temperature, using equation (15). Finally, for 1,000 draws of the outcome, we estimate (1) and plot the resulting estimates and distributions of the β_k. See Figure 5.

The regression estimates imply positive effects of exposure to hot and cold temperature bins, despite the fact that there is no actual relationship between contemporaneous temperature and the outcome (see (15)). Looking across the many temperature bins, binning bias becomes increasingly severe as we move toward the temperature extremes. While Figure 5a presents a U-shaped bias, reversing the sign of the trend in the outcome leads to inverse-U-shaped bias (Figure 5b), and removing the differential trend in the outcome eliminates the bias (Figure 5c).

Intuitively, the magnitude of the spurious effects depends not just on the magnitude of differential trends induced in the temperature variables through binning, but also on the magnitude of the outcome trends. This can be seen formally in the theoretical treatment, where the bias increases with ω_Y (see Result 2). To illustrate this dimension of the binning bias problem with real temperature data and many temperature bins, we further simulate outcomes using the following model:

$$Y_{ct} = \beta \cdot t \cdot Temp_c^{Base} + \xi_{ct} \tag{16}$$

where $\xi_{ct} \sim \mathcal{N}(0, \sigma)$ and where σ is the standard deviation of $t \cdot Temp_c^{Base}$. We vary β—governing the magnitude of outcome trends—across specifications, letting β range

from 0.0001σ to 0.01σ. Coefficient estimates across all values for β are reported in Figure A.1a. Analogous estimates for negative trends in the outcome are reported in Figure A.1b. U-shapes and inverted U-shapes become more pronounced for higher β.

4.3 Real temperature data and calibrated outcomes trends

To gauge whether binning bias could be quantitatively meaningful in realistic settings, we calibrate the trend in our simulated outcomes to those of variables that have been studied in the context of climate damage estimates. To do so, we estimate:

$$Y_{ct} = \delta_c + \eta_t + \alpha_c \cdot t + \epsilon_{ct} \tag{17}$$

where Y_{ct} is actual outcome data and $\hat{\alpha}_c$ captures any county-specific time trend for that outcome. After estimating county-specific trends using equation (17) for each outcome variable, we run the following model:

$$\hat{\alpha}_c = \omega + \zeta \cdot Temp_c^{Base} + \epsilon_c \tag{18}$$

where $\hat{\zeta}$ captures any correlation between baseline temperatures and county-specific time trends. For example, Figure A.2 plots the cross-county relationship between $Temp_c^{Base}$ and $\hat{\alpha}_c$, when the outcome variable in equation (17) is the county-level annual mortality rate for people age 65 and over. We estimate a positive and significant relationship, indicating that mortality is declining relatively more rapidly in colder places.

Finally, we simulate Y_{ct} so that its trend in baseline temperature matches each outcome of interest:

$$\widetilde{Y}_{ct} = \hat{\delta}_c + [\hat{\omega} + \hat{\zeta} \cdot Temp_c^{Base}] \cdot t + \xi_{ct} \tag{19}$$

where $\hat{\delta}_c$ is estimated from equation (17) and $\hat{\omega}$ and $\hat{\zeta}$ are estimated from equation (18).

Using these calibrated outcomes, together with the real temperature data, we then estimate the standard panel model (1). By construction in (19), there is no actual contemporaneous relationship between temperature realizations and outcomes here, just differential trends, so an unbiased model should estimate $\beta_k = 0$ for all temperature bins.

We begin with the outcome calibrated to trends in mortality for people over 65 years old. Figure A.3a reports the estimates: we find large, positive effects of both extreme heat and extreme cold on the outcome calibrated to trends in mortality despite there being no actual effect of contemporaneous temperature on the outcome in our simulation. That

said, our results to this point do not imply that existing estimates are necessarily biased; instead, they suggest that a standard panel model does not distinguish between a causal effect of temperature and a spurious effect driven by differential trends.

The other panels in Figure A.3 report estimates of equation (1) in which the trends in $\widetilde{Y}_{ct}$ are calibrated to other real world outcomes of interest, including crime rates by type of crime and crop yields. In all cases simulations using these calibrated outcomes lead to spurious U-shaped or inverse-U-shaped effects. That said, the magnitude of the effect varies substantially across outcomes. The spurious effects are largest in magnitude when the outcome is calibrated to trends in mortality; they are somewhat smaller when the outcome is calibrated to trends in violent and non-violent crime; and they are very close to zero when calibrated to trends in corn yields, since trends in corn yields have only a weak relationship with baseline temperatures. These differences suggest that binning bias could affect regression estimates very differently depending on the differential trend in the outcome variable. This intuition is confirmed when we evaluate the actual extent of binning bias in these outcomes, implementing various potential solutions to the binning bias problem, in Section 6.

4.4 Heterogeneous effects and "adaptation"

There is growing interest in understanding potential mitigating effects of human adaptation. A common strategy for studying adaptation estimates the effects of extreme heat separately for already-hot locations, which have plausibly learned to adapt, and for colder locations, which have not (e.g., Butler and Huybers, 2013; Carleton et al., 2022; Hultgren et al., 2022). Weaker marginal effects of hot days in hotter locations are then interpreted as adaptation. However, the extent of binning bias–and the temperature bins in which the bias is most severe–may differ systematically between hotter and colder places (or more generally, between any two sub-samples). This has the potential to generate spurious results that look like "adaptation" but reflect different differential trend parameters.

Recall from the bias formulas (equations (12) and (13)) that bias depends on trends with baseline temperature (the ω parameters) and noise in these trends (the σ^2 parameters). These relationships will change when dividing the data into hot and cold subsamples, and we can examine these changes explicitly in the data. Specifically, we estimate county-specific trends for cold days (days below 10°F) and hot days (days above 90°F) and relate these trends to baseline temperature separately for counties with below

and above median baseline temperature. Figure 6a plots these relationships for trends in cold days. Cold places (shown in blue) have a steeper slope in baseline temperature (higher ω_C) but also a much higher variance (higher $\sigma^2_{e_C}$), compared to the hot places (shown in red). Figure 6b shows that the converse is true for hot days, where hot places show the steeper slope and higher variance when relating county-specific trends to baseline temperature.

Now consider the resulting binning bias within these separate hot and cold county samples. To gain intuition, we can substitute the estimated trend and variance parameters from Figures 6a and 6b into the bias formulas (equations (12) and (13)).[13] The bias estimates, along with their components, are reported in Table A.2. For the cold bin, the formula produces substantially larger bias in hot counties (28.58) compared to cold ones (0.72). Interestingly, the larger bias among hot counties is not due to a steeper relationship between cold day trends and initial temperature;[14] instead, it is driven by by substantially lower noise in these trends ($\sigma^{hot}_{e_C} << \sigma^{cold}_{e_C}$; see Figure 6a). This is a case where trends having a small but precise relationship to initial temperature can drive especially large bias. Intuitively, precisely ordered, differential trends across the hot counties means that smaller variations in cold day counts can now predict the given variation in the outcome (i.e., generate a larger regression coefficient; see also the discussion following Result 2). Meanwhile, for the hot bin, the bias estimate is larger for cold counties (6.41) compared to hot ones (1.60), driven by substantially lower trend variance among the cold counties and following similar logic (see Table A.2).

Overall, these are the patterns that might lead a researcher to conclude there is "adaptation" – hot temperatures have a larger effect in cold places, and cold temperatures have a larger effect in hot places. However, this pattern is just an artifact of differential trends in the binned temperature variables. Of course, the model in Section 2 and resulting bias formula are stylized, with just three bins (cold, moderate, hot). What happens in practice, with many temperature bins?

To investigate this, we return to the simulations of Section 4.2 and report estimates of equation (1) separately for counties with above versus below median values of $Temp_c^{Base}$. The relationship of the outcome variable trend to baseline temperature is taken to be the same in both sets of places. The results are reported in Figure 6c. We find a larger spurious "effect" of extreme cold in hotter counties and, to a lesser extent, a larger "effect"

[13] We simulate the outcome variable so that $\omega_Y = 1$.

[14] The relationship with initial temperature is actually less steep among hot counties ($\omega_C^{hot} < \omega_C^{cold}$).

of extreme heat in colder counties.[15] That is, we see spurious results that tend to look like adaptation—the effect of extreme temperatures appears weaker in places when these temperatures are more common hence 'adapted'—but it is driven by differences in the extent of binning bias.

4.5 Sensitivity and Extensions

We further consider alternatives to our baseline simulation analysis, exploring other non-linear specifications of the temperature damage function, other units of aggregation, and other temperature datasets.

Alternative non-linear temperature damage specifications. As a non-parametric method, binning the data approximates non-linear representations of temperature variables. "Binning bias" will then intuitively extend to other non-linear methods, and is not just about the specific binning approach per se. Here we consider several alternatives.

First, while our baseline specification uses bins that are ten degrees Fahrenheit wide, we alternatively consider finer bins (five degrees wide) and coarser bins (twenty degrees wide). See Figures 7a and 7b. Binning bias appears robustly.

Second, we consider "killing degree days" (KDDs), which counts the number of degree days in excess of some temperature threshold. It is commonly used when studying the exposure of crop species to extreme heat (e.g., Schlenker and Roberts, 2009; Butler and Huybers, 2013). Figure 7c shows that if we replace binned temperature in our simulation analysis with the number of KDDs in excess of 80, 85, 90, or 95° F, we obtain a spurious positive effect, even when there is no actual effect of KDDs on the outcome.

Third, we consider polynomials in daily temperature realizations. While most estimates of the non-linear effects of temperature rely on the binned specification outlined in Section 2, some instead use a polynomial of daily temperatures, summed across the year (e.g., Carleton et al., 2022). One motivation is to reduce the number of estimated parameters (e.g., four when using a fourth-order polynomial compared to nine in our binned specification). We conduct a version of the simulation from Section 4.2 using the polynomial parameterization instead of bins and find similar levels of bias to our baseline results (Figure 7d).

[15] Alternative specifications further show and can strengthen the apparent (but spurious) adaptation result - see Figure A.4a and A.4b.

Together, these results show that binning bias is a general issue for estimates of the non-linear effects of temperature. Bias emerges for any bin size, for alternative weighting schemes across daily realizations (e.g., killing degree days), and for alternative approaches to aggregating daily temperature data (e.g., polynomials).

Other units of aggregation. So far, our results have focused on county-year panels in the U.S.. A large share of existing work on climate change and climate adaptation relies on this structure, but in other cases, researchers aggregate, either spatially (e.g., to states) or temporally (e.g., to decades rather than years). Binning bias remains an issue in both cases. Specifically, we show that we obtain a very similar bias when we instead use U.S. states as the unit of analysis instead of counties (Figure A.5a) or when we instead aggregate to the county-decade level (Figure A.5b).

Alternative temperature data. In our main analysis, we use temperature data from the ERA-5 database. However, similar results hold using other temperature data sets, including U.S. county-level temperature data from Schlenker (2024), which uses a balanced panel of weather stations from the PRISM database (Figure A.5c), and using raw station level data from GHCN (Menne et al., 2012) (Figure A.5d).[16] We find similar levels of bias performing our simulations using both alternative temperature datasets.

5 Solutions to Binning Bias

In this section, we investigate potential empirical solutions to binning bias. In Section 5.1, we develop a method to directly control for the place-specific temperature dynamics that generate binning bias. In Section 5.2, we consider reduced-form approaches. Across these varied methods, we assess which have the potential to ameliorate binning bias, and which do not appear to do so.

For each potential solution, we run the panel model following the simulation methods in Section 4. Specifically, we use the real temperature data and 1,000 draws of the outcome variable according to (15). We then report the bin coefficient estimates when implementing the candidate solution. If the correction is successful, these coefficient

[16] We aggregate these data to the county level following the approach in Deschênes and Greenstone (2011), which takes the inverse of the squared distance to the county centroid as the weights for each station within a 200km radius.

estimates should become centered around zero for every temperature bin. We also zoom into the top (over 90° F) and bottom (below 10° F) bins and display the full distribution of coefficient estimates and corresponding t-statistics across all 1,000 simulated runs. This presentation allows us to see the resulting shape of the coefficient distribution and the share of coefficient estimates that would spuriously reject a null effect. Finally, for each potential solution, we also report the model-derived bias estimate and the components of the bias formula (Table A.1). This makes it possible to directly compare any remaining bias across potential solutions, as well as identify which components of the bias formula are affected by each approach.

5.1 Controlling for non-random temperature exposure

Our first strategy seeks to control explicitly for the non-random component of temperature exposure in each temperature bin and time period. To do so, we build a statistical model of the counterfactual temperature distribution in each time and place. Using this counterfactual distribution, the researcher can calculate the expected value of each temperature bin variable and then add these as controls to the panel regression. Once one controls for the non-random portion of each temperature bin in each time and place, the identifying variation that is left is the random component of the realization—the idiosyncratic shocks that motivated the panel approach in the first place.

This approach builds on work by Wooldridge (2015) and Borusyak and Hull (2023), which show that controlling directly for the expected value of a non-random treatment variable can remove the bias.[17] This approach does not require knowing anything about the functional form of the dependent variable, is more parsimonious econometrically than the reduced-form approaches presented below, and could in theory be applied in any context. The main challenge, however, is correctly estimating counterfactual temperature distributions that can be used to construct the control variables.

To fix ideas, suppose that we know the distribution of temperature days in each location-year; i.e., the distribution $F_{ct}(x)$, where x is a given temperature. From this distribution, we can calculate the expected number of days in each temperature bin. In particular, consider a bin $k(x, x')$ that counts the number of days in the temperature

[17] This is also related to the idea of a "propensity score" (Rosenbaum and Rubin, 1983).

range between x and x'. The expected number of days in bin k is:

$$\widehat{T}_{k(x,x'),ct} = 365 \times (F_{ct}(x') - F_{ct}(x)) \tag{20}$$

For example, for the bin from 85 to 90 degrees, the expected number of days in that bin in a given county-year is given by $365 \times (F_{ct}(90) - F_{ct}(85))$. As shown above, this expected day count will typically be on a faster upward trend for hotter counties, introducing bias.

The key point is that the stochastically realized number of days in each bin, T_{kct}, is random conditional on controlling for the expected number of days in each bin, $\widehat{T}_{kct}$. Thus, we can estimate the augmented model that includes non-random exposure to each temperature bin as a control:

$$Y_{ct} = \alpha_c + \lambda_t + \sum_k \beta_k T_{kct} + \sum_k \xi_k \widehat{T}_{kct} + u_{ct} \tag{21}$$

The fact that rising temperatures cause differential trends in different temperature bins is now accounted for by $\widehat{T}_{kct}$. Thus, the β_k coefficients from estimates of equation (21) are unbiased.

Simulations with a known temperature DGP. To validate this approach, we first return to the simulation of Section 4.1 in which we directly specify the data generating process for temperature. Since the independent variables are constructed from simulated data with a known normal distribution, for each county-year we can compute the expected number of days that fall into each temperature bin using (20). We then include the expected number of days in each bin as controls in the panel model, as in equation (21). If the approach in equation (21) is valid, estimates of each β_k should be equal to zero.

Implementing this approach, we find indeed that controlling for the expected number of days in each bin removes the bias (Figure A.6). The method is successful both when the trend in the outcome is a linear function of baseline temperature (Figure A.6a) or when we use more complicated, non-linear functional forms for both the outcome and temperature variables (Figure A.6b). Moreover, when we add in *actual* time-specific effects (e.g., where extra days in the hottest and coldest bin do affect outcomes), we recover those effects accurately (Figure A.6c). This suggests that this approach can work well, and does not require specifying the correct functional form of the trend in the outcome.

The challenge, then, is applying this strategy to real-world temperature data, when

the counterfactual distribution is not known. We turn to this next.

Constructing counterfactual temperature distributions. To apply the method above with real-world temperature data, we build a statistical model of changes in expected (non-random) exposure to each temperature bin. Our statistical model of temperature realizations allows for flexible cross-sectional and temporal differences in temperature distributions. Specifically, we allow temperature distributions, $F_{ct}(x)$, to be arbitrarily different in different counties c to capture the fact that some places have a higher temperature variance in a given year than others. We also allow temporal dynamics to be arbitrarily different in each county to capture the fact that warming has not been uniform across locations. Moreover, to allow for different temporal dynamics by season, we let the warming process evolve separately for each county-month.[18]

Our counterfactual procedure is as follows:[19]

1. Define x_{cdmt} as the observed temperature for county c, day d, month m and year t.
2. Construct the mean observed temperature, $\bar{x}_{cmt}$, for each county-month-year.
3. Estimate the linear time trend in mean temperature for each county-month across all years using the following regression:

$$\bar{x}_{cmt} = \mu_{cm} + \gamma_{cm} \cdot t + \epsilon_{cmt} \tag{22}$$

   where $\hat{\gamma}_{cm}$ is the estimated linear trend coefficient for month m in county c (e.g., the trend in mean January temperature in Suffolk county). In our main analysis, we use empirical Bayesian methods to shrink our estimates of the γ_{cm} toward the average estimate, though we show results with and without this step.[20]

[18] An alternative approach to building a statistical model could be to use a climate model to estimate counterfactual temperature distributions. There is some existing work in atmospheric science and climatology on approximating local temperature distributions; however, these "emulators" tend to not perform as well for extreme temperatures, which are particularly important for our applications (e.g., Wang et al., 2024) and building a climate model that could be incorporated into our analysis would be a non-trivial task, given the state of the literature. For example, we implemented the emulator-based temperature distribution functions from Wang et al. 2024 in our context using the approach in equation (21), and they were not able to remove binning bias. That said, this could be a promising area of future work.

[19] We provide code to create these counterfactuals using any daily temperature dataset using the *cftemp* Stata package, which will be available on our website.

[20] We construct a posterior value for the trend which is a weighted average of the actual estimate and the average estimate across all counties, where the weight corresponds to the ratio between the variance of all estimated coefficients and the sum of this variance and the estimated coefficient variance.

4. Return to the actual daily temperature realizations, x_{cdmt}, and subtract off the county-month mean shift that has taken place up to time t, as estimated in (22):

$$x_{cdmt}^{detrend} = x_{cdmt} - \hat{\gamma}_{cm} \cdot t \tag{23}$$

5. Combine all de-trended temperature realizations in each county-month into a single distribution. Denote this distribution $F_{cm}(x)$. For example, with 31 days in January, and 50 years of data, the distribution for January in Suffolk county will include $31 \cdot 50 = 1550$ de-trended temperature realizations. This allows us to use the full set of data to flexibly capture the underlying temperature distribution in each county-month.
6. To identify the potential empirical distribution for each county, month, *and year*, we project the county-month distributions forward using the county-month specific time trend $\hat{\gamma}_{cm}$. This generates an empirical distribution of potential temperature realizations for each county, month, and year, denoted $F_{cmt}(x)$.
7. Finally, we aggregate $F_{cmt}(x)$ across months within each county-year to generate the counterfactual temperature distribution at the county-year level, $F_{ct}(x)$. We can then use equation (20) to estimate the share of the distribution in each county-year that falls into each temperature bin. After normalizing by the number of days in the year, this generates our estimate of non-random exposure, $\widehat{T}_{kct}$.

As an alternative approach, rather than estimate a separate trend for each county-month in equation (22), we can use a county-specific fourth-order Chebychev polynomial to approximate seasonal variation in temperature (as in Schlenker and Roberts, 2006) and estimate the place-specific time trend for each of the four Chebychev polynomial terms.[21] This has the advantage of requiring a smaller number of estimated parameters (four per county instead of twelve per county) and allows temperature to vary smoothly over the year instead of changing discontinuously in each month. In practice, however, this approach performs similarly to the one outlined above.

[21] Specifically, let the normalized calendar day variable be $x \in [-1, 1]$, so that January 1 is -1 and December 31 is 1. Then the nth order Chebyshev polynomial is defined recursively as $T_0(x) = 1, T_1(x) = x$, and $T_n(x) = 2x \cdot T_{n-1}(x) - T_{n-2}(x)$ for $n \geq 2$. For each county separately, we regress daily temperatures on these polynomials up to the fourth order and their respective interactions with the year variable, estimating jointly with the constraint that the predicted temperatures on January 1 equals that of December 31. For identification, we omit December 31 from the regression. To construct our counterfactual temperature distribution, we take the coefficients on these interaction terms and use them similarly to the $\hat{\gamma}_{cm}$ in the process described above, i.e., "de-trend" the temperature realizations to create a single empirical distribution as in step 4, and project the distribution forward as in step 6.

Correcting for binning bias. Once we have the expected day counts, $\widehat{T}_{kct}$, in hand, we can return to our simulations from Section 4.2—which used real temperature data—and see whether controlling for $\widehat{T}_{kct}$ eliminates binning bias.

Adding $\widehat{T}_{kct}$ as controls, we find that the coefficient estimates are now all centered at zero (see Figure 8a). Since there is no true effect of any temperature bin on the outcome, this indicates that controlling for the non-random component of exposure to each bin removed binning bias. Moreover, when there *are* causal effects of contemporaneous temperature on the outcome, this approach accurately recovers the true effect (see Figure 8b). Figure 9 plots the full coefficient and t-statistic distributions across simulations for the top and bottom temperature bins. When we control for the counterfactual temperature estimates, the extreme temperature coefficient estimates are centered around zero and reject the null hypothesis at the appropriate rate. The results for our baseline approach are shown in green; however, the results are similar if we exclude the trend shrinkage step (orange in Figure 9) or use Chebychev polynomials (brown in Figure 9).

If we examine the bias formula in Table A.1, we find that these approaches reduce coefficient bias *not* by reducing the residual trend in the outcome, which could change across applications, but rather by fixing the binning bias on the "right hand side" of the panel regression—by correcting for differential trends in exposure to extreme temperature. This indicates that this approach could be applicable across a broad range of outcome variables and contexts.

5.2 Reduced form strategies

There are a range of potential reduced-form refinements of the main regression specification of interest (equation (1)), including specifications that include additional controls. To the extent that additional controls successfully absorb trends in the right- or left-hand-side variables that drive the bias, they could reduce the extent of binning bias and also be easier to implement than the approach outlined in the previous section.

State-by-year fixed effects. First, studies occasionally include state-by-year fixed effects when analyzing the county-level effects of temperature. While there is debate whether including such fixed effects is best practice given that it absorbs much of the variation in temperature (see, e.g., Fisher et al., 2012), we investigate whether their inclusion solves binning bias. While the bias is somewhat attenuated, state-by-year fixed effects do not eliminate spurious effects of temperature in simulations (see Figure 10a).

The bottom and top bins falsely reject zero (p <0.05) in 14% and 26% of the simulation runs, respectively (see Figure 11; results with state-by-year fixed effects are shown in orange). Furthermore, the reduction in false-rejections comes primarily from the fact that state-year fixed effects increase noise, rather than reduce the mean bias in the coefficients, and a large share of coefficient estimates are far from zero. Thus, while state-by-year effects may reduce bias in cases where there is substantially less within-state variation in outcome or bin trends, it does not appear to be a generally useful solution.

Lags of dependent and independent variables. Second, we investigate the inclusion of lags of the dependent and independent variables as additional controls. Existing work has shown that appropriately accounting for the relationship between contemporaneous temperature and its lags is important for isolating the exogenous component of temperature shocks in cross-country growth regressions (Nath et al., 2024). However, we find that in this case accounting flexibly for the lag structure of both temperature and the outcome, by allowing for 5 lags of both Y and X variables, does not address binning bias (see Figure 10b). The coefficients on the top and bottom temperature bins both falsely reject zero over 90% of the time (Figure 11).[22]

Place-specific trends. Third, we consider the inclusion of location-specific linear trends. For example, in a county-year regression, one could include a separate linear trend for each county. This approach appears in some studies, though most papers do not include place-specific trends.[23] We find that the inclusion of county-specific linear trends solves the problem in our baseline simulation (see Figures 10c and 11), but this is precisely because in our simulations the differential trend in the outcome is itself linear. Thus, adding a linear county-specific trend as a control fully captures the trends in the outcome model. This can be seen clearly in Table A.1, which shows that the including county-specific trends fully removes residual trends in the outcome (column 5).

The broader issue, however, is that if one does not know the functional form of the outcome trend, it is not clear that place-specific *linear* trends will work. For example, if we conduct a modified version of our simulation in which the trends in the outcome are

[22]Intuitively, lags can help account for common dynamics in a panel but, like general time fixed effects, are not well suited to dealing with differential, place-specific dynamics.

[23]Papers that implement place-specific trends include, for example, Barreca et al. (2016), who study mortality at the U.S. state level and include quadratic trends at the state-month level, and Burke et al. (2015), who study growth at the country level and include country-specific quadratic trends.

quadratic rather than linear, the inclusion of place-specific linear trends does not perform as well (see Figure A.7).[24] The same is true for a cubic outcome trend (see Figure A.9).

Of course, a possible rejoinder to this is that adding quadratic or cubic place-specific trends as a control would in turn solve this problem—and indeed, if the trends in the outcome were in fact quadratic or cubic, this is true. However, there are two challenges to this approach. First, it only works if the researcher is confident that they know the functional form of the trend in the outcome variable. While in our simulations we generated the trends ourselves, in real-world applications the functional form of the trends is unknown. Second, it comes with a substantial degrees-of-freedom cost. In a county-by-year level regression, for example, adding county-specific linear trends requires adding over 3,000 controls; place-specific higher order trends would add commensurately more. This can reduce power in many practical applications. So while including place-specific trends may be useful in some applications, it may not work in general.

Isolating short-panel comparisons. Fourth, motivated by the model, we consider the inclusion of interactions between place and broader time-period fixed effects; e.g. including county-by-five-year-interval fixed effects. This specification only allows for comparisons among years in a given location within the same narrow time window, within which trends in temperature or the outcome may be less stark. As T grows, the time trend in temperature swamps the idiosyncratic variation in temperature; hence, conversely, by restricting the analysis to comparisons within 5 year bins in a given place, the time trends will be very small compared to the idiosyncratic variation, and so the bias will be limited.

While we are unaware of prior work that implements this approach, we find that it entirely eliminates the bias in our baseline simulation (see Figure 10d). The simulated coefficient estimates on the top and bottom bin are centered around zero and reject the null at the appropriate rate (see Figure 11). This approach also performs well with more complicated—quadratic and cubic—trends in the outcome (see Figures A.7 and A.9).

This approach seems to successfully eliminate binning bias in our simulations. This is consistent with the theoretical finding that longer panels make binning bias worse (see Result 2). However, there are two potential drawbacks. First, it remains *ad hoc* and

[24] We simulate outcome variables as $\widetilde{Y}_{c,t} = t^2 \cdot T_c^{Base} + \epsilon_{c,t}$ where $\epsilon_{c,t}$ is drawn from a normal distribution centered at zero and with common variance. We simulate temperature variables as $\widetilde{\mu}_{ct} = Temp_c^{Base} + (\frac{9}{250})^2 \cdot (t-1)^2$ where $t \in \{1, 50\}$. For each year and county, we draw 365 daily temperatures from a normal distribution centered at $\widetilde{\mu}_{ct}$ with a common fixed variance and assign these to their corresponding temperature bins.

it is not clear how effective it is as a general approach. This is especially true since it seems to have reduced overall bias by eliminating the residual trend in the outcome (see Table A.1), rather than the differential trend in baseline temperature (though, unlike county-specific linear trends, it does so also for the more complicated outcome trends). Second, this approach is only relevant in applications that use high-frequency (e.g., yearly) temperature variation since it requires exploiting variation *within* slightly aggregated time windows. This would not be applicable in situations when the unit of observation is a location-by-decade (for example) or when using long-difference specifications. Some outcome data do not exist at an annual frequency (e.g., the U.S. Population Census, as we explore below, or the U.S. Census of Agriculture data, as in Moscona and Sastry, 2023). Moreover, aggregating across time is a common strategy to understand the potential role of human adaptation (e.g., Burke and Emerick, 2016) or simply to compare the effects of long-run and short-run exposure to temperature extremes. Thus, this approach may not be relevant for some contexts.

6 Applications

The final step of our analysis is to investigate how correcting for binning bias affects estimates of the relationship between temperature bins and a variety of real-world outcomes. We focus on four sets of outcomes that are relevant for understanding the economic consequences of climate change: county-level population, mortality, crop yields, and crime.

For consistency, we use the same specifications, temperature data (ERA5-land; see Section 3.1) and variable construction throughout.[25] For each outcome, we first estimate a naive specification, which includes counts of days in each temperature bin as well as county and time fixed effects (equation (1)). We then estimate augmented versions that include our suggested counterfactual controls for the expected number of days in each bin (equation (21)). The results are shown in Figure 12. We also perform similar exercises with the two reduced-form approaches that showed the most promise, county-specific linear trends, shown in Appendix Figure A.11, and (where possible) the 'short-panel'

[25]Different papers in the literature use different temperature datasets, and many use slight variants on the specification in equation (1), including several of the approaches described in Section 5.2, such as state-year fixed effects, lags, and county-specific trends. Replicating each paper's exact specification and exact temperature dataset for the hundreds (if not thousands) of papers in this literature is beyond the scope of this paper. Rather, our goal here is to highlight the importance of ensuring binning bias is addressed moving forward. We provide the code used to construct the counterfactual temperature distributions so that other authors can use them in whatever context is of interest.

approach using five-year bin indicators interacted with county fixed effects, shown in Appendix Figure A.12.

Population We first examine total county-level population. The reallocation of population, including through migration, has long been considered an important potential mechanism of adaptation to environmental change (e.g., McLeman and Smit, 2006; Cruz and Rossi-Hansberg, 2024). We estimate the effect of exposure to extreme temperatures on the log of county-level population in each decade from the U.S. census.

We report the effect of temperature extremes on log population, both with and without our proposed solution to binning bias (Figure 12a). The blue points display the β_k estimates from the panel specification that does not control for non-random exposure to each temperature bin. These estimates show a U-shape, and suggest that exposure to both extreme cold and extreme heat is *positively* associated with increased population. However, when we include the controls to correct for binning bias (red points), the estimates suggest a much flatter relationship between temperature and population, with the exception of a significant, negative effect of the top temperature bin. Controlling for county-specific linear trends also results in a substantially flatter relationship (Figure A.11a). Including aggregated time fixed effects interacted with county fixed effects is not possible in this context.[26]

Thus, the naive and bias-corrected estimates paint very different pictures of how population patterns respond to temperature fluctuations. The former suggest that population moves *toward* parts of the country with greater exposure to temperature extremes. This could be driven by the fact that there has been an average movement toward warmer parts of the U.S., like the Sun Belt and Florida, during the past decades, and these warm regions also have steep positive trends in the upper temperature bins. When we account for binning bias we find a much flatter relationship. Further, including counterfactual temperature controls suggests that exposure to extreme heat causes population declines, consistent with adaptation to extreme heat by moving away from the most heat-exposed locations (Figure 12a).

[26]The unit of observation for population is a county-decade, making it impossible to exploit variation within each decade or shorter periods. Hence it is not possible to deploy the county-by-five-year fixed effect approach above. This illustrates the limitations of this approach for applications where outcome data are not reported annually.

Mortality The effect of temperature extremes on mortality has been studied extensively (e.g., Deschênes and Greenstone, 2011; Barreca et al., 2016; Carleton et al., 2022) and the estimated positive effect of extreme heat on mortality represents a large share of predicted economic damages from climate change (Hsiang et al., 2017). We explore how binning bias affects estimates of the relationship between extreme heat and mortality, which we measure as the mortality rate (deaths per 100,000) for people age 65 and over (the population shown to be most sensitive to extreme heat in prior work).

While naive estimates of the effect of temperature on mortality return the U-shaped relationship that is familiar from the literature, where both high and low temperature extremes increase mortality rates, we find much weaker evidence of this pattern when we account for binning bias by controlling for the non-random component of exposure to each bin (Figure 12b). The effects of temperature extremes are substantially smaller in magnitude and (if anything) only marginally significantly different from the omitted bin (even though the standard errors are smaller after controlling for non-random exposure). In this particular specification, for example, the magnitude of the effects of the coldest bins (below 10°F and 10°-20°F) fall by more than half once we control for the expected temperature realizations, and the effect in the highest bin (>90°F) falls to 0. Both reduced form methods to account for binning bias tell a very similar story (Figures A.11b and A.12b), and the pattern is similar if we pool effects on mortality across all age groups (see Figures A.13a, A.14a, and A.15a).[27] These results suggest that the estimates of the effect of extreme temperatures on mortality may be affected by binning bias.

Crop Yields The agricultural sector is perhaps most exposed to environmental change, and a large literature has established a stark, negative effect of extreme heat on crop yields (e.g., Schlenker and Roberts, 2009; Hatfield and Prueger, 2015; Hultgren et al., 2022). We estimate the effect of temperature on corn yields and find a clear, inverted-U shaped relationship both in the native specification and in the specification that controls for non-random exposure to each temperature bin (Figure 12c). We continue to find the same relationship after employing both reduced-form strategies to correct for binning bias (Figures A.11c and A.12c). We replicate the same analysis for wheat and soybean yields

[27] This slightly more complicated pooled age-group specification is based on the empirical approach in existing work used to take into account differing mortality trends and impacts across age groups (e.g., Deschênes and Greenstone, 2011). The pooled regression is the same specification as equation (1) but with separate observations for each age group (ages < 1, 1-44, 45-64, and > 65), weighted by the population of each age group-county-year.

and again find little evidence that binning bias affects the estimates (Figures A.13b and A.13c). Thus, estimates of the relationship between crop yields and temperature do not seem biased (in either direction) by spurious trends, perhaps because yield trends across countries are not strongly correlated with average temperature.

Crime Weather patterns have been shown to affect both non-violent and violent crime (e.g., Brunsdon et al., 2009; Ranson, 2014; Cohen and Gonzalez, 2024). Studies propose different mechanisms—ranging from changes in the opportunity cost of criminal activity to direct psychological effects of high temperatures—but most find a positive relationship between extreme heat and violence. To examine the relationship between temperature and crime, we estimate an augmented version of the baseline panel model (equation (1)) in which, following the convention in the literature, the unit of observation is the county-year-month and we include county, month and year fixed effects. We construct analogous counterfactual measures of non-random exposure to each temperature bin at the county-month level instead of the county-year level (see e.g., Ranson, 2014). We also modify our reduced-form strategies by including county-month specific linear trends (instead of county-specific linear trends) and interacting five-year bin fixed effects with each county-month pair (instead of each county).

Naive estimates of the relationship between temperature and violent crime suggest that violent crime rate is higher at high temperatures. There appears to be a negative effect (relative to the omitted group) of cold days, but no significant effect of extremely cold days ($< 10°$F). When we control for non-random exposure to each bin, the effect of heat on crime is smaller for the extremely hot temperature bin while substantially stronger in the cold bins (Figure 12d). The reduced-form approaches also tell a similar story (Figures A.11d and A.12d). In this case, the correction both moderates one conclusion from the naive estimates (i.e., where extreme heat leads to more violent crime) and strengthens what had been a weaker relationship (i.e., that cold reduces crime).

Turning to non-violent crime, the naive estimates suggest that exposure to cold days substantially reduces crime rates while exposure to the hottest days increases crime rates. Once we correct for binning bias, the negative effect of cold days on non-violent crime is even larger in magnitude while the positive effect of hot days disappears (Figure A.13d).

7 Conclusion

A large literature uses panel models to estimate temperature's effects on various economic outcomes. To investigate potential non-linear effects of temperature, existing work often considers exposure to distinct temperature ranges—very cold days, moderately cold days, moderately hot days, extremely hot days, and so on. A common finding is U-shaped or inverted U-shaped relationships, where more extreme temperatures have larger effects. These estimates have become important not only for informing the impacts of temperature, but also for quantifying potential damages from global climate change and computing the "social cost of carbon" that can guide policy intervention.

In this paper, we explore how warming itself may introduce econometric challenges when estimating non-linear temperature effects. We show that warming mechanically introduces place-specific trends in extreme temperature bins (i.e., counts of hot and cold days) that are strongly correlated with the place's initial temperature. If trends in outcome variables also correlate with baseline temperatures for any reason, then panel models can generate spurious U-shaped relationships between temperatures and outcomes, similar to those seen in existing literature. Using theory, simulations, and real-world data, we show when and how this bias can occur and argue that it may be relevant across important applications.

We then turn to potential solutions. We develop a "counterfactual" approach that controls for the expected number of days in each temperature bin in each place and time, and we show how to feasibly calculate this counterfactual distribution using empirical temperature realizations. In our simulations, this approach eliminates the bias and uncovers the true effects of temperature. If the counterfactual temperature distribution is not available, including place-specific linear trends or place-by-period fixed effects also appear to substantially ameliorate the problem; however, these approaches are applicable in a more limited set of contexts. Other off-the-shelf reduced-form solutions, including richer time effects (e.g., state-by-year effects) or including lags of the dependent and independent variables, do little to solve binning bias.

Finally, we apply our proposed corrections to a range of real empirical outcomes, including population counts, mortality, crop yields, and crime. In some cases, results become substantially weaker when the correction is applied. In other cases, the bias appears to have masked important results. And in other cases, results seem unaffected. Overall, addressing and correcting for binning bias appears important for obtaining reliable

estimates of the impact of temperature.

Future work may develop additional counterfactual methods and corrective strategies and apply them in broader settings. We have focused here on candidate applications using U.S. data, but applying and evaluating corrections in different spatial and temporal settings across the globe, and for additional outcome variables, present important areas for further research.

References

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

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Brunsdon, C., Corcoran, J., Higgs, G., and Ware, A. (2009). The influence of weather on local geographical patterns of police calls for service. *Environment and Planning B: Planning and Design*, 36(5):906–926.

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

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

Butler, E. E. and Huybers, P. (2013). Adaptation of us maize to temperature variations. *Nature Climate Change*, 3(1):68–72.

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Kopp, R. E., McCusker, K. E., Nath, I., et al. (2022). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Quarterly Journal of Economics*, 137(4):2037–2105.

Cohen, F. and Dechezleprêtre, A. (2022). Mortality, temperature, and public health provision: evidence from mexico. *American Economic Journal: Economic Policy*, 14(2):161–192.

Cohen, F. and Gonzalez, F. (2024). Understanding the link between temperature and crime. *American Economic Journal: Economic Policy*, 16(2):480–514.

Cruz, J.-L. and Rossi-Hansberg, E. (2024). The economic geography of global warming. *Review of Economic Studies*, 91(2):899–939.

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

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):740–798.

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

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

EPA (2023). Report on the social cost of greenhouse gases: Estimates incorporating recent scientific advances. Technical report, U.S. Environmental Protection Agency. Accessed: 2024-03-18.

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. *American Economic Review*, 102(7):3749–3760.

Graff Zivin, J. and Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1):1–26.

Hatfield, J. L. and Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and climate extremes*, 10:4–10.

Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D. J., Muir-Wood, R., Wilson, P., Oppenheimer, M., et al. (2017). Estimating economic damage from climate change in the united states. *Science*, 356(6345):1362–1369.

Hultgren, A., Carleton, T., Delgado, M., Gergel, D. R., Greenstone, M., Houser, T., Hsiang, S., Jina, A., Kopp, R. E., Malevich, S. B., et al. (2022). Estimating global impacts to agriculture from climate change accounting for adaptation. *Available at SSRN 4222020*.

McLeman, R. and Smit, B. (2006). Migration as an adaptation to climate change. *Climatic change*, 76(1):31–53.

Menne, M. J., Durre, I., Korzeniewski, B., McNeill, S., Thomas, K., Yin, X., Anthony,

S., Ray, R., Vose, R. S., Gleason, B. E., and Houston, T. G. (2012). Global historical climatology network - daily (ghcn-daily), version 3.32. [Accessed 2025-04-12].

Moscona, J. and Sastry, K. A. (2023). Does directed innovation mitigate climate damage? evidence from us agriculture. *The Quarterly Journal of Economics*, 138(2):637–701.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9):4349–4383.

Nath, I. B., Ramey, V. A., and Klenow, P. J. (2024). How much will global warming cool global growth? Technical report, National Bureau of Economic Research.

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

Ranson, M. (2014). Crime, weather, and climate change. *Journal of environmental economics and management*, 67(3):274–302.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Schlenker, W. (2024). Daily weather data - continental USA (version 2018).

Schlenker, W. and Roberts, M. J. (2006). Nonlinear effects of weather on corn yields. *Review of agricultural economics*, 28(3):391–398.

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

Wang, M., Souza, A. N., Ferrari, R., and Sapsis, T. (2024). Stochastic emulators of spatially resolved extreme temperatures of earth system models. *Authorea Preprints*.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

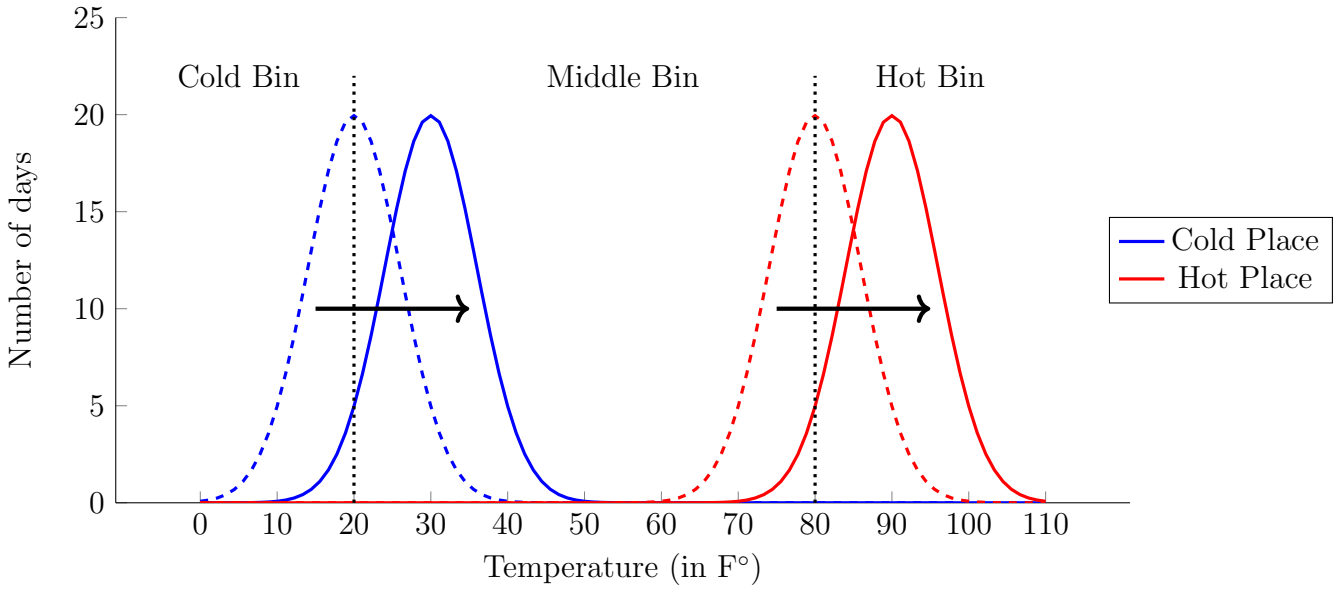


Figure 1: Temperature Bins with Uniform Warming

Notes: This figure depicts the impact of rising temperatures on exposure to different temperature ranges for a hypothetical "cold" location and a hypothetical "hot" location. The vertical lines delineate different temperature bins, where days are counted in a cold bin, middle bin, and hot bin, respectively. With warming, the number of hot days rises in the hot place but not in the cold place. Similarly, the number of cold days declines in the cold place, but not in the hot place. The hot place will thus experience differential upward trends in counts of both hot days and cold days, compared to the cold place.

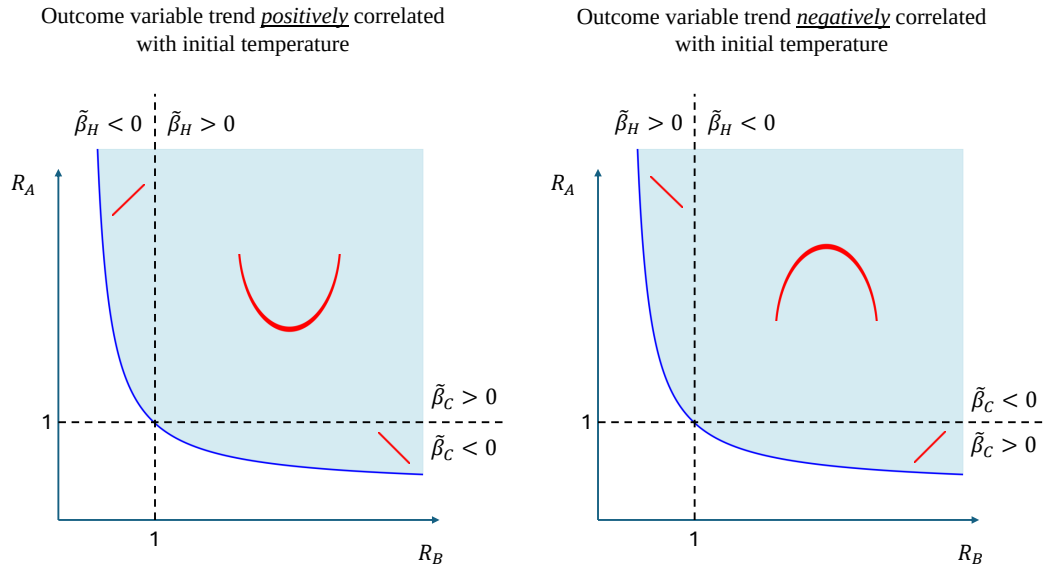


Figure 2: The Direction of Bias

Notes: This figure considers the sign of the biases on the coefficients $\hat{\beta}_C$ and $\hat{\beta}_H$ using Result 1. U-shapes and inverted U-shapes are common, but upward and downward sloping biases are also possible, with the bias directions indicated in red. The vertical and horizontal axes are ratios formed from the numerators of the bias equations. The ratio $R_A = \text{Var}(\delta_{H,i})\text{Cov}(\delta_{C,i}, \gamma_i)/[\text{Cov}(\delta_{C,i}, \delta_{H,i})\text{Cov}(\delta_{H,i}, \gamma_i)]$ is drawn from the bias equation (7). The ratio $R_B = \text{Var}(\delta_{C,i})\text{Cov}(\delta_{H,i}, \gamma_i)/[\text{Cov}(\delta_{C,i}, \delta_{H,i})\text{Cov}(\delta_{C,i}, \gamma_i)]$ is drawn from the bias equation (8). The condition $R_A R_B > 1$ is necessary and guarantees a non-singular regression matrix. In the left panel, which features U shapes, we assume that the outcome trends are positively covariant with the trends in both hot and cold days. In the right panel, which features inverted U shapes, we assume that the outcome trends are negatively covariant with the trends in both hot and cold days.

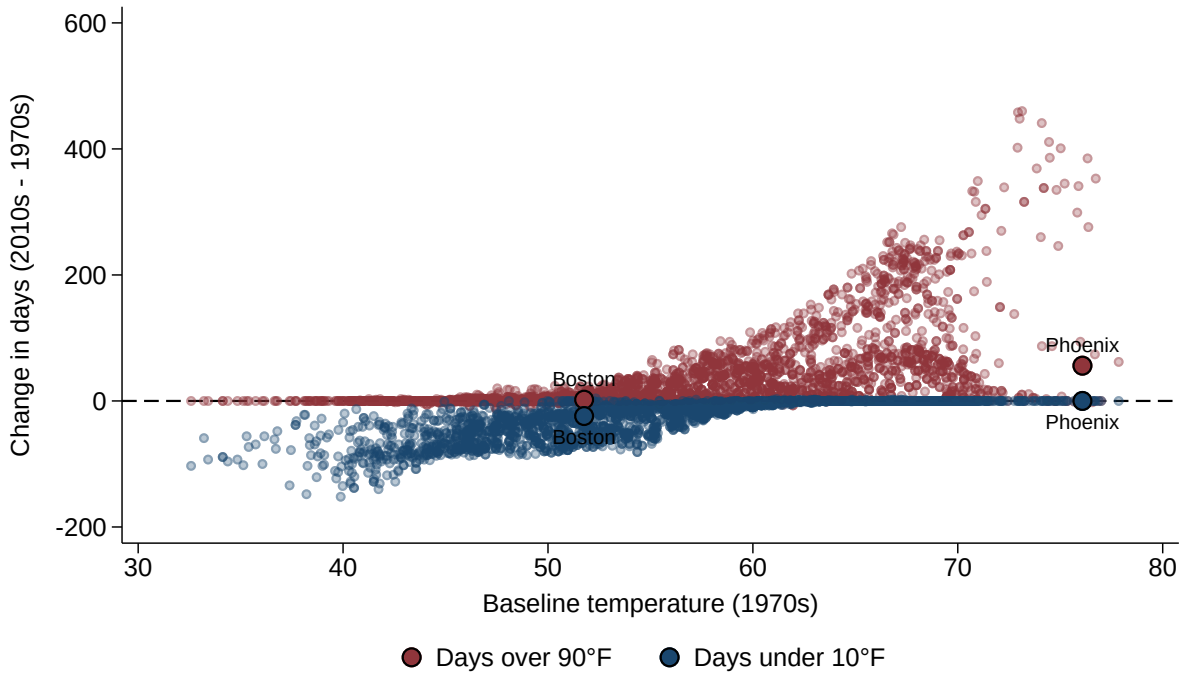
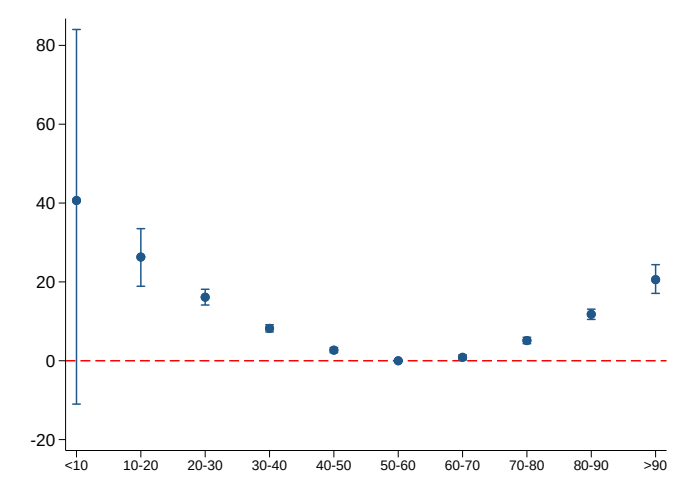
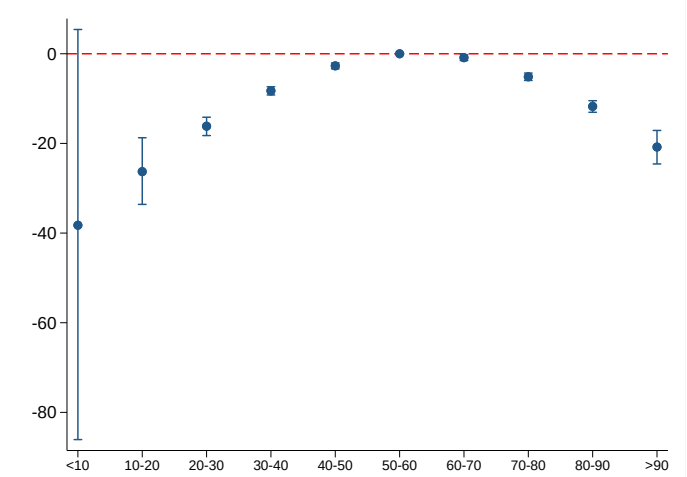


Figure 3: Baseline Temperature and Trends in Temperature Extremes

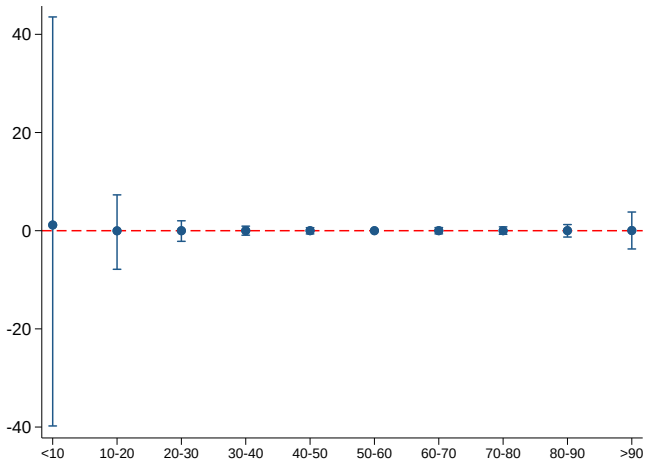
Notes: This figure plots the U.S. cross-county relationship between average temperature during the 1970s and changes in exposure to cold ($< 10°$F) and hot ($> 90°$F) days between the 1970s and 2010s. Temperature is measured using the ERA-5 data set. Mean temperature on each day in each county is computed as the average temperature between 7am and 6pm in the ERA-5 grid cell closest to the county centroid. The relationship between baseline temperature and the change in the number of cold days in each county is shown in blue, and the relationship between baseline temperature and the change in the number of hot days in each county is shown in red. The counties that contain Boston and Phoenix are highlighted in both scatter plots. Relative to initially cold places, initially hot places exhibit relative upward trends in counts of extremely hot days and extremely cold days.



(a) Positively Correlated Outcome Trend



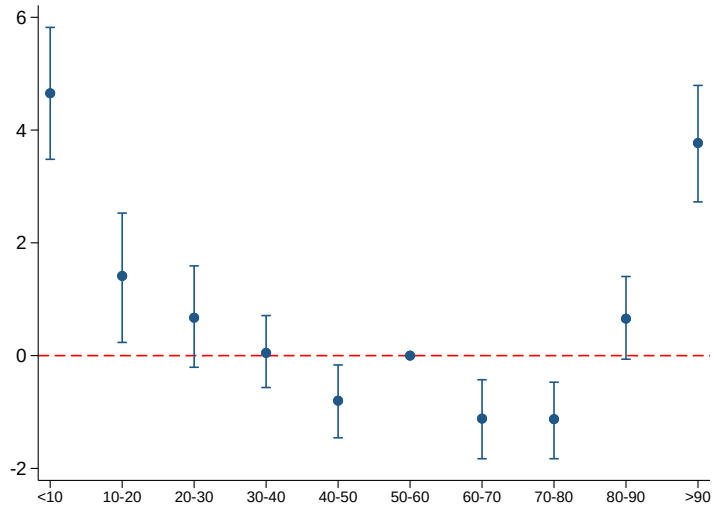
(b) Negatively Correlated Outcome Trend



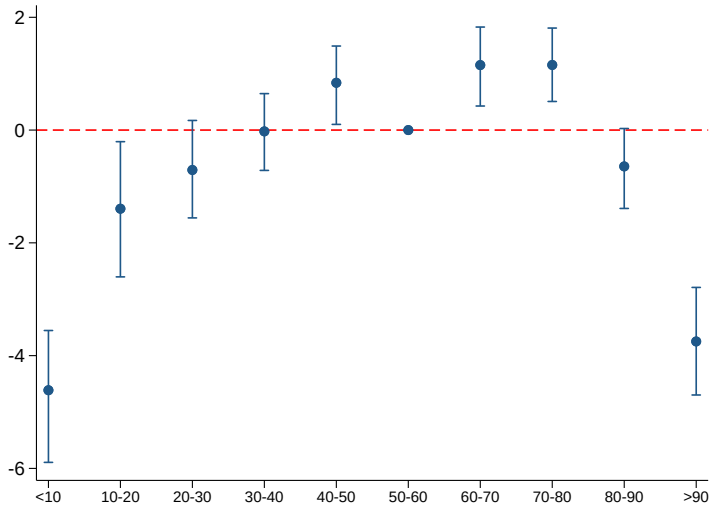
(c) Uncorrelated Outcome Trend

Figure 4: Binned Regression Estimates with Simulated Outcome and Uniform Temperature Change

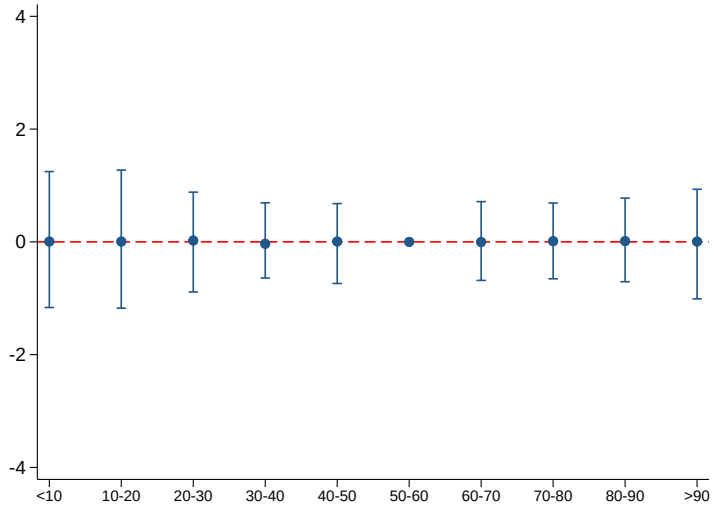
Notes: Each sub-figure reports estimates from the panel model (1), in which the outcome is simulated using the following equation: $Y_{ct} = aTemp_c^{base} + \epsilon_{ct}$. The error term is drawn from a normal distribution centered at zero with variance equal to the variance of $aTemp_c^{base}$ over the full sample. The temperature data are drawn so that all counties have the same warming trend (1.8°F over the sample period), starting with a normal distribution centered at the county-level baseline temperature. In each sub-figure, we report the mean and 95% confidence intervals of each coefficient from (1) taking 1,000 separate draws of the outcome and temperature variables. In Panel (a), the outcome is simulated to have a positive differential trend in baseline temperature (i.e., $a = t$); in Panel (b), the outcome is simulated to have a negative differential trend in baseline temperature (i.e., $a = -t$); and in Panel (c) the outcome has no relationship with baseline temperature (i.e., $a = 0$).



(a) Positively Correlated Outcome Trend



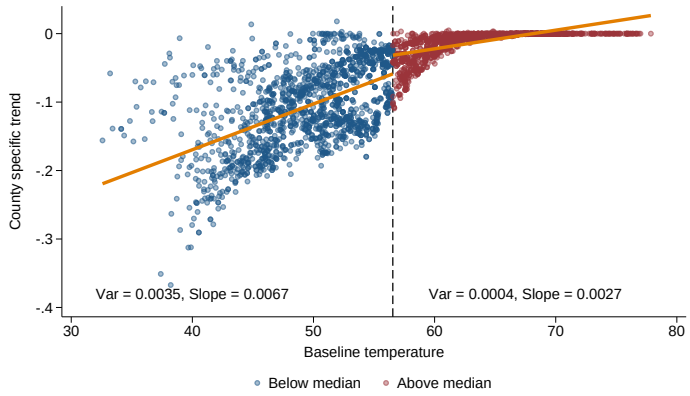
(b) Negatively Correlated Outcome Trend



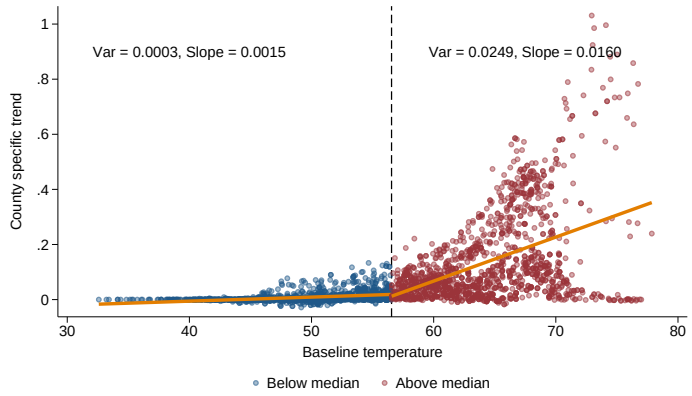
(c) Uncorrelated Outcome Trend

Figure 5: Binned Regression Estimates with Real Temperature Data and Dynamics

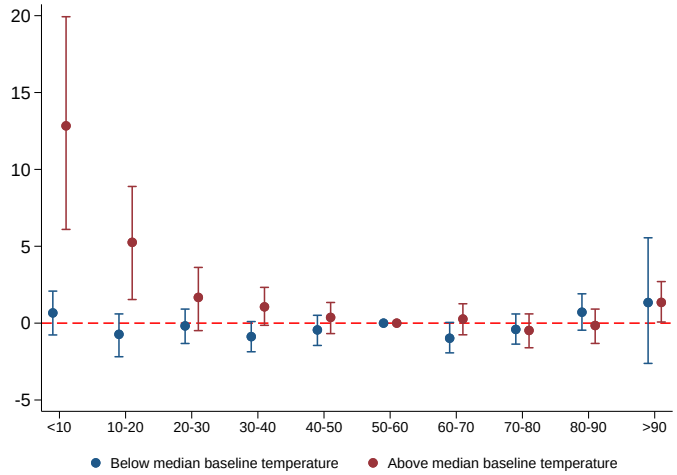
Notes: Each sub-figure reports estimates from equation (1), in which the outcome is simulated using the following equation: $Y_{ct} = aTemp_c^{base} + \epsilon_{ct}$. The error term is drawn from a normal distribution centered at zero with variance equal to the variance of $aTemp_c^{base}$ over the full sample. The temperature bins are constructed from historical, daily temperature realizations using the ERA-5 Land dataset. In each sub-figure, we report the mean and 95% confidence intervals of each coefficient from (1) taking 1,000 separate draws of the outcome variable. In Panel (a), the outcome is simulated to have a positive differential trend in baseline temperature (i.e., $a = t$); in Panel (b), the outcome is simulated to have a negative differential trend in baseline temperature (i.e., $a = -t$); and in Panel (c) the outcome has no relationship with baseline temperature (i.e., $a = 0$).



(a) County-specific Trends in Cold Days



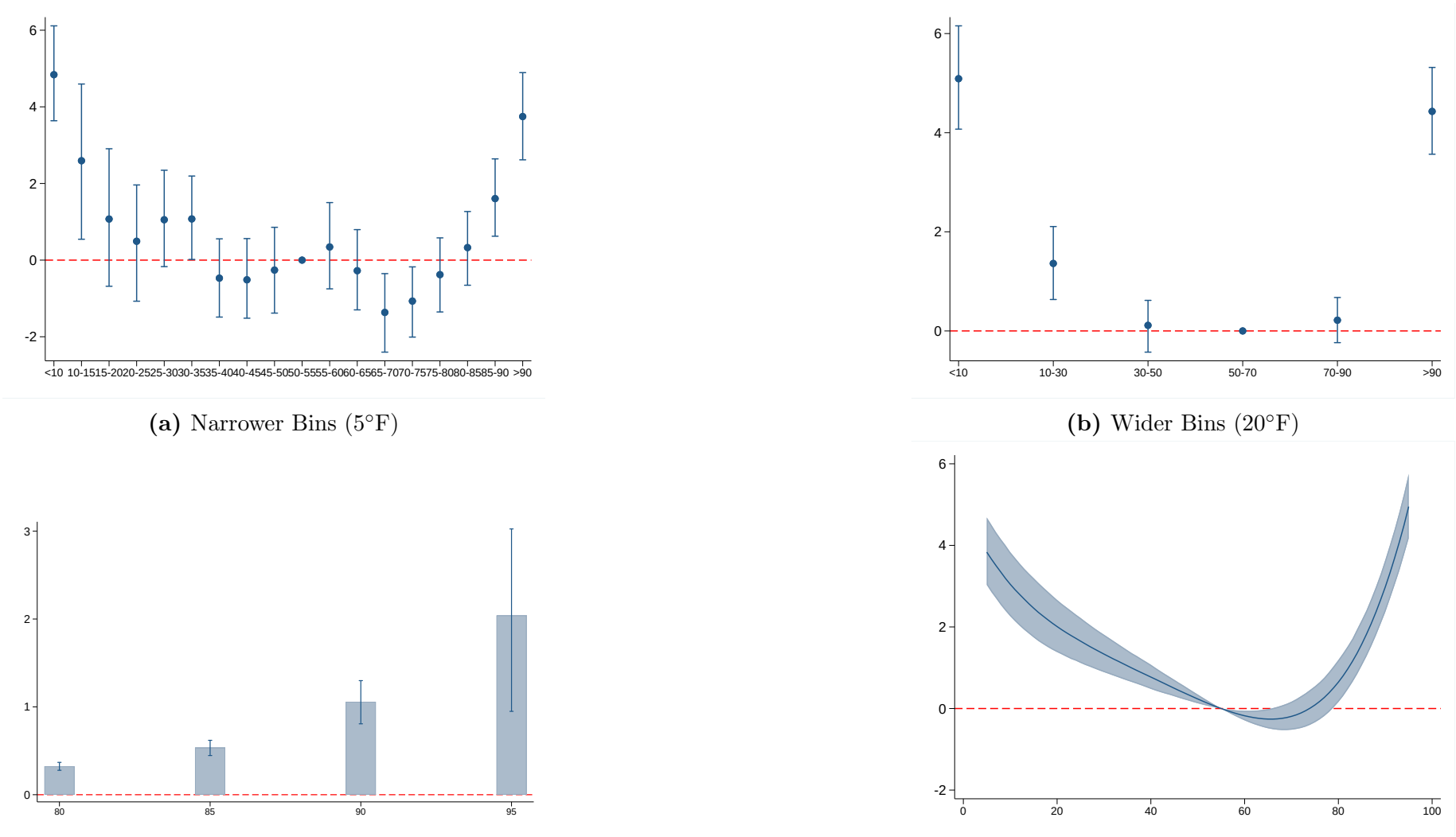
(b) County-specific Trends in Hot Days



(c) Binned Regression Estimates

Figure 6: Binned Regression Estimates and "Adaptation"

Notes: Panel (a) reports the relationship between baseline temperature and county-specific trends (1970s-2010s) in the number of days below 10°F. We report this relationship separately for countries with above and below median average temperatures, and report the slope and trend variance for both sets of counties. Panel (b) reports the relationship between baseline temperature and county-specific trends (1970s-2010s) in the number of days above 90°F. We report this relationship separately for countries with above and below median average temperatures, and report the slope and trend variance for both sets of counties. Panel (c) reports estimates of equation (1), separately for counties with above versus below median baseline temperature. The outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. We report the mean and 95% confidence intervals for each coefficient estimate, taking 1,000 draws of ϵ_{ct} from a normal distribution centered at zero with variance equivalent to the in-sample variance of $tTemp_c^{base}$.



(a) Narrower Bins (5°F)

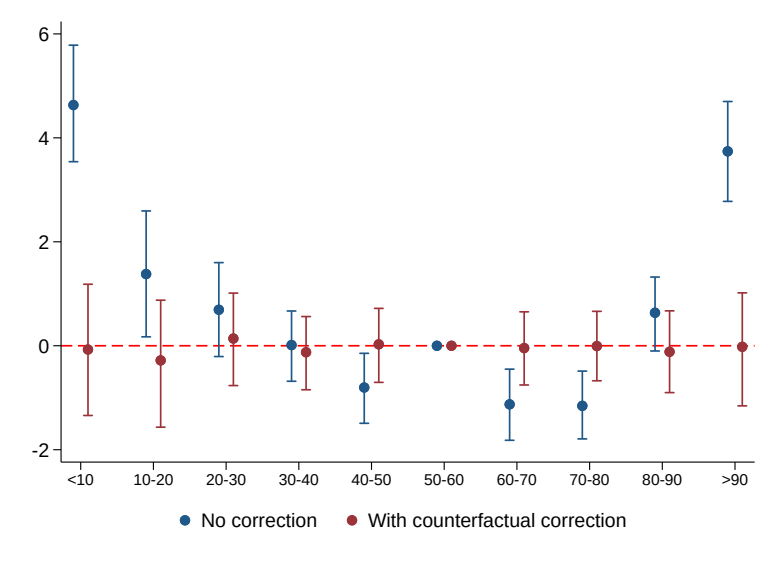
(b) Wider Bins (20°F)

(c) Killing Degree Days

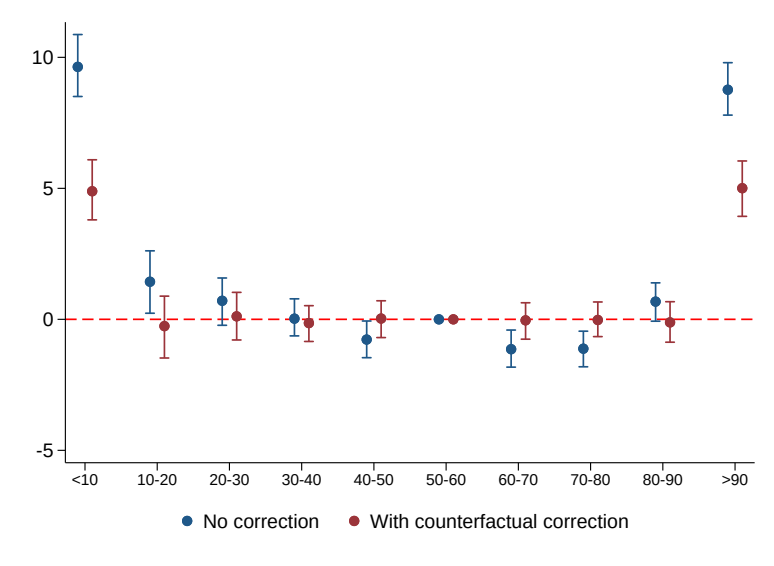
(d) Polynomials in Daily Temperature

Figure 7: Binning Bias: Alternative Nonlinear Specifications

Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data, under a variety of alternative non-linear parameterizations of temperature shocks. Panels (a) and (b) present results when daily temperatures are binned at 5°F and 20°F intervals, respectively. Panel (c) implements a regression of the simulated outcome on killing degree days (KDDs). For a given threshold T^*, KDD is defined as $\sum_{t=T^*}\sum_{d=1}\mathbb{1}(t < Temp_d \leq t+1)](t+0.5-T^*)$ (i.e., the number of degree-days in a given year above a threshold T^*). We report the mean and 95% confidence intervals for the coefficient estimates on KDD from setting 80, 85, 90, and 95°F as the threshold. Panel (d) present results from the polynomial specification used in Carleton et al. (2022). This specification regresses the outcome variable on a fourth-order polynomial of daily temperatures with county fixed effects and year fixed effects. Each polynomial term is constructed by raising the daily temperatures to that power and summing them within each county-year. This figure plots the predicted values of the fourth-order polynomials at each °F, where we normalize by subtracting the predicted value of 55°F. In all specifications, the outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. We report the mean and 95% confidence intervals for each estimate, taking 1,000 draws of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to the in-sample variance of $tTemp_c^{base}$.



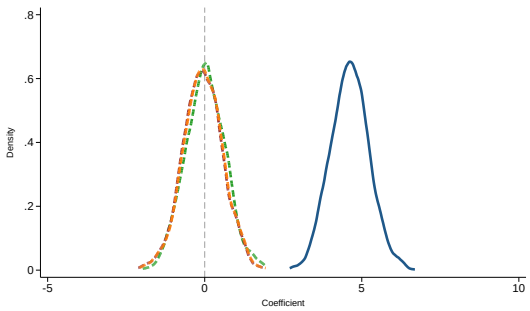
(a) Counterfactual Control Correction, No True Effect



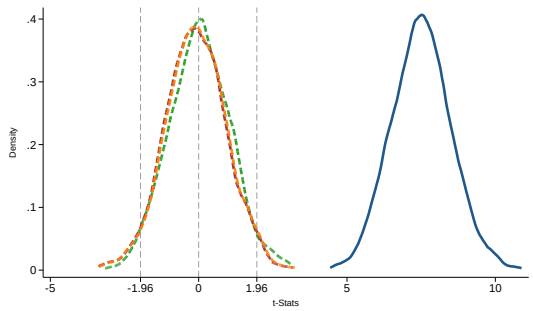
(b) Counterfactual Control Correction, True Effect of Extreme Temperatures

Figure 8: Correcting for Binning Bias with Counterfactual Temperature Controls

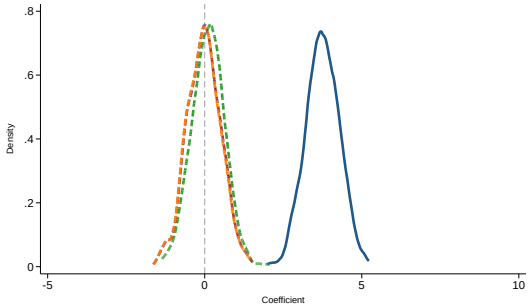
Notes: Both sub-figures report estimates of equation (1) alongside estimates of equation (21) that also include the expected counterfactual exposure to each temperature bin as controls. In Panel (a), the outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) but no causal effect of exposure to any temperature bin. In Panel (b), the outcome has the same differential trend but there is also a positive effect of 5 of exposure to the top and bottom temperature bins. Daily temperature realizations used to construct each temperature bin are from the ERA-5 Land dataset. For each regression specification, we report the mean and 95% confidence intervals for each coefficient estimate, taking 1,000 draws of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to the in-sample variance of $tTemp_c^{base}$.



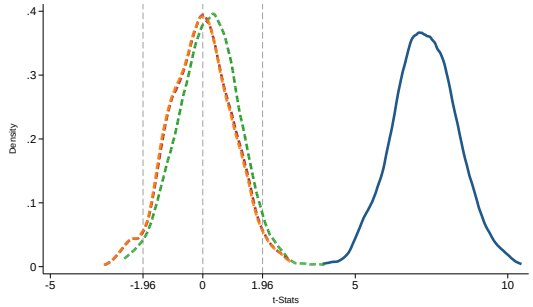
(a) Under 10 °F Bin: Coefficient



(b) Under 10 °F Bin: t-Stat



(c) Over 90 °F Bin: Coefficient



(d) Over 90 °F Bin: t-Stat

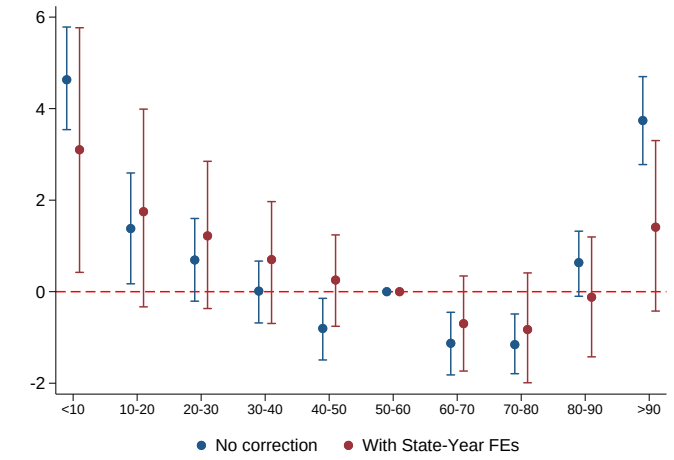
Figure 9: Comparison of Counterfactual Temperature Solutions with Simulated Outcome Variable (Linear Trends)

—— No correction (Under 10 Bin: 100.0%, Over 90 Bin: 100.0%)

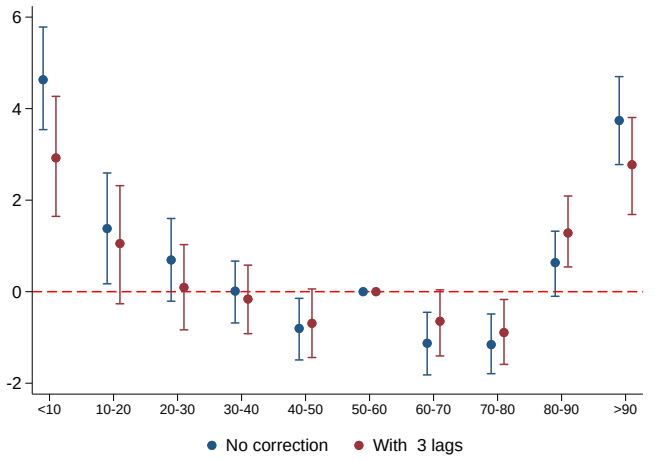
- - - Counterfactual Temperature Control (5.4%, 5.5%)

- - - Counterfactual Temperature Control without Bayes (8.8%, 8.2%) - - - Chebyshev (8.4%, 4.9%)

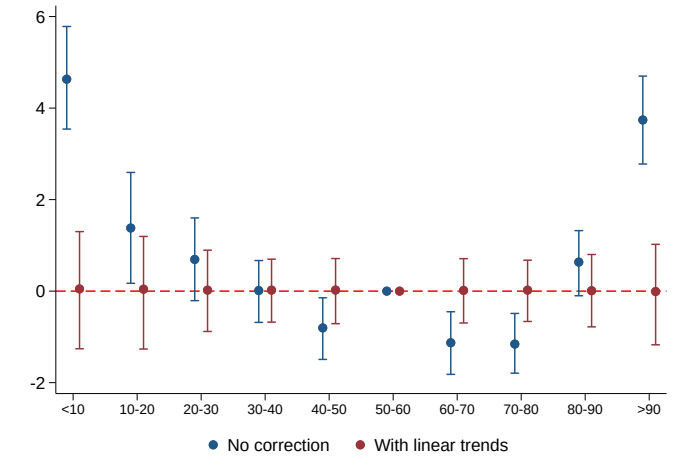
Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data. Here, we plot the density of coefficients and t-statistics from estimating equation (5) 1,000 times, which only uses the two extreme temperature bins, alongside densities from specifications that also include the expected counterfactual exposure to each temperature bin as controls. Each color and line plots densities that use different methods to construct the counterfactuals, as indicated in the legend. The outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. Each time, we take a draw of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to two-times the in-sample standard deviation of $tTemp_c^{base}$. Numbers inside the parentheses in the legend denote the number of runs per 1,000 simulations that the coefficient is significant at the 5% level, for the under 10°F bin and the over 90°F bin respectively.



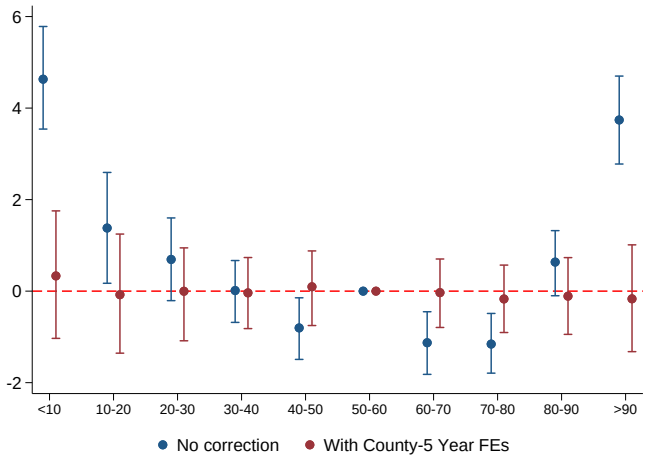
(a) State-Year Fixed Effects



(b) Lags of Dependent/Independent Variables



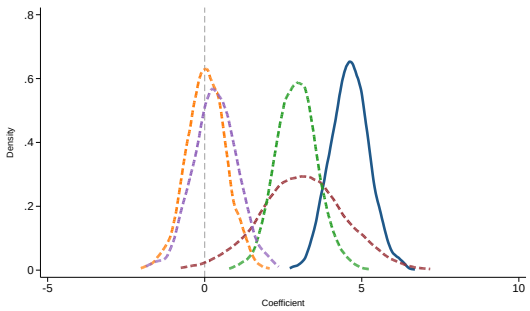
(c) County-Specific Linear Trends



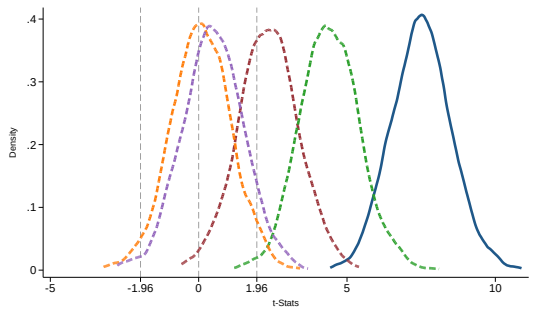
(d) Five-Year Indicators × County Fixed Effects

Figure 10: Correcting for Binning Bias with Reduced-Form Approaches

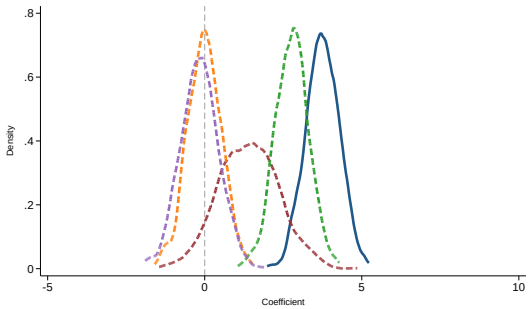
Notes: Simulations are the same as in Figure 8, but here we report estimates of equation (1) alongside estimates that employ alternative, reduced-form corrective strategies as indicated in the title for each panel. Daily temperature realizations used to construct each temperature bin are from the ERA-5 Land dataset. The outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$). For each regression specification, we report the mean and 95% confidence intervals for each coefficient estimate, taking 1,000 draws of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to the in-sample variance of $tTemp_c^{base}$.



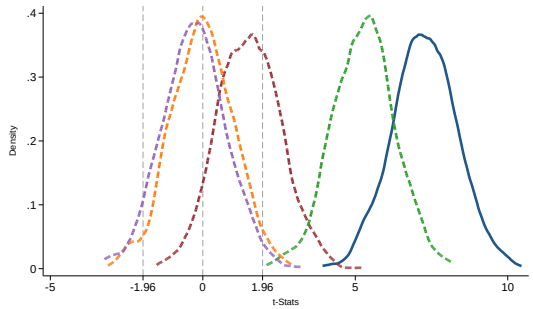
(a) Under 10 °F Bin: Coefficient



(b) Under 10 °F Bin: t-Stat



(c) Over 90 °F Bin: Coefficient



(d) Over 90 °F Bin: t-Stat

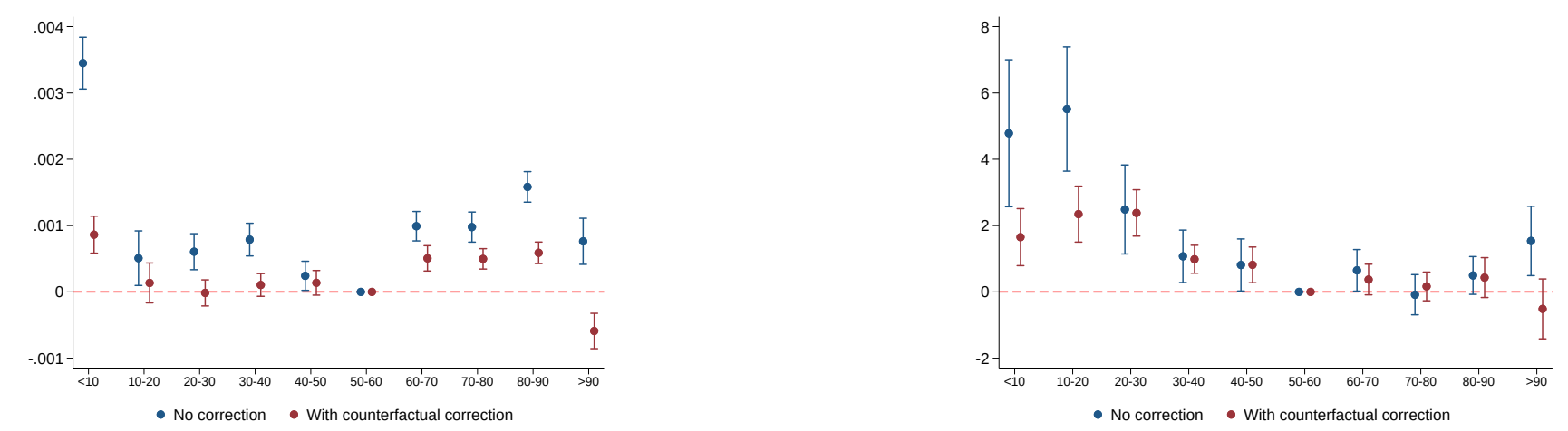
Figure 11: Comparison of Reduced Form Solutions with Simulated Outcome Variable (Linear Trends)

—— No correction (Under 10 Bin: 100.0%, Over 90 Bin: 100.0%) - - - State-Year Fixed Effects (41.4%, 72.8%)

- - - With 3 Lags (100.0%, 100.0%) - - - County-Specific Linear Trends (5.4%, 4.5%)

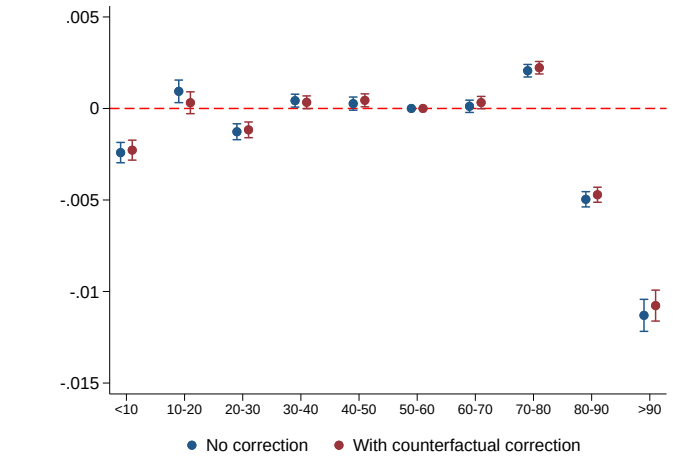
- - - County-5 Year Fixed Effects (10.8%, 5.2%)

Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data. Here, we plot the density of coefficients and t-statistics from estimating equation (5) 1,000 times, which only uses the two extreme temperature bins, alongside densities from specifications that use alternative reduced-form corrective strategies as indicated in the legend. The outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. Each time, we take a draw of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to two times the in-sample standard deviation of $tTemp_c^{base}$. Numbers inside the parentheses in the legend denote the number of runs per 1,000 simulations that the coefficient is significant at the 5% level, for the under 10°F bin and the over 90°F bin respectively.

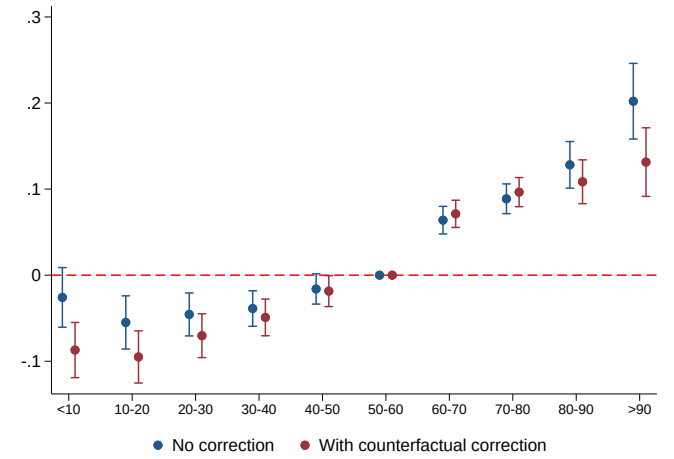


(a) log Population

(b) Over-65 Mortality Rate



(c) log Corn Yield



(d) Violent Crime Rate

Figure 12: Binning Bias with Real Outcome Data (Counterfactual Control Correction)

Notes: This figure reports estimates of (1) for a series of outcome variables, alongside estimate of (21) that also include the expected counterfactual exposure to each temperature bin as controls. Daily temperature realizations in each county used to construct each temperature bins are from the ERA-5 Land dataset. In Panel (a), the unit of observation is a county-decade and the outcome is log of population measured from the U.S. Census (1970-2010). In Panel (b), the unit of observation is a county-year (1970-2002) and the outcome variable is the mortality rate for individuals over the age of 65 (per 100,000 inhabitants), measured using the CDC mortality files. In Panel (c), the unit of observation is a county-year (1985-2006) and the outcome variable is log of corn yield, measured using NASS Quickstats. In Panel (d), the unit of observation is a county-month (1970-2009) and the outcome variable is violent crime rate (i.e., murder, rape, and aggravated assault) measured using data from the FBI. 95% confidence intervals are reported and standard errors are clustered by county.

	Boston		**Phoenix**	
	1970s	2010s	1970s	2010s
Days with daytime mean below 10°F	20	10	0	0
Days with daytime mean above 90°F	0	0	1,000	1062

Table 1: Changes in Extreme Temperature Exposure in Boston and Phoenix

Notes: This table reports the number of days below 10°F and above 90°F in Boston and Phoenix, during both the 1970s and 2010s. Daily temperature realizations are from the ERA-5 Land dataset and the daily temperature is defined as the average temperature between 7am and 6pm.

Appendix

A Proofs

A.1 Proof of Result 1

We first remove place and time fixed effects by applying the two-way fixed effects transformation. We can then proceed to the estimated treatment coefficients from the residualized regression specification, following the Frisch–Waugh–Lovell theorem. For the true model, the adjusted (demeaned) outcome equation is:

$$\tilde{Y}_{it} = \beta_C \tilde{C}_{it} + \beta_H \tilde{H}_{it} + \tilde{\gamma}_i \tilde{t} + \tilde{\varepsilon}_{it} \tag{24}$$

Residual cold days are

$$\tilde{C}_{it} = \tilde{\delta}_{C,i} \tilde{t} + \tilde{v}_{C,it} \tag{25}$$

where $\tilde{\delta}_{C,i} = \delta_{C,i} - \delta_{C,\text{avg}}$ (and $\delta_{C,\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} \delta_{C,i}$) and where $\tilde{t} = t - \bar{t}$ (and $\bar{t} = \frac{1}{T} \sum_{s=1}^{T} s$).

Residual hot days are

$$\tilde{H}_{it} = \tilde{\delta}_{H,i} \tilde{t} + \tilde{v}_{H,it} \tag{26}$$

where $\tilde{\delta}_{H,i} = \delta_{H,i} - \delta_{H,\text{avg}}$ (and $\delta_{H,\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} \delta_{H,i}$).

Residual place-specific trends in the outcome variable are

$$\tilde{\gamma}_i = \gamma_i - \gamma_{\text{avg}} \tag{27}$$

where γ_{avg} is the average place-specific trend across all places, $\gamma_{\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} \gamma_i$.

Thus, the two-way fixed effects eliminate unit intercepts and the common time pattern but leave unit-specific deviations.

A.1.1 Estimating the Model without Individual-Specific Trends

When we estimate the model without accounting for the residual individual-specific trend $\tilde{\gamma}_i \tilde{t}$, we have:

$$\tilde{Y}_{it} = \beta_C \tilde{C}_{it} + \beta_H \tilde{H}_{it} + \tilde{u}_{it} \tag{28}$$

where the error term is:

$$\tilde{u}_{it} = \tilde{\gamma}_i \tilde{t} + \tilde{\varepsilon}_{it} \tag{29}$$

raising potential for bias if the residual trending terms in either $\tilde{C}_{it}$ or $\tilde{H}_{it}$ have any covariance with the residual trend in the error term $\tilde{u}_{it}$.

A.1.2 OLS Estimators

The OLS estimators for β_C and β_H are:

$$\begin{pmatrix} \hat{\beta}_C \\ \hat{\beta}_H \end{pmatrix} = \left(\tilde{\mathbf{D}}'\tilde{\mathbf{D}}\right)^{-1} \tilde{\mathbf{D}}'\tilde{Y} \tag{30}$$

where $\tilde{\mathbf{D}}$ is the matrix of adjusted treatment variables:

$$\tilde{\mathbf{D}} = \begin{pmatrix} \tilde{C}_{11} & \tilde{H}_{11} \\ \tilde{C}_{12} & \tilde{H}_{12} \\ \vdots & \vdots \\ \tilde{C}_{NT} & \tilde{H}_{NT} \end{pmatrix} \tag{31}$$

A.1.3 Bias in the Estimators

The bias in the estimated coefficients arises due to the correlation between $\tilde{u}_{it}$ and the treatment variables $\tilde{C}_{it}$ and $\tilde{H}_{it}$.

$$\begin{pmatrix} \text{Bias}(\hat{\beta}_C) \\ \text{Bias}(\hat{\beta}_H) \end{pmatrix} = \left(\tilde{\mathbf{D}}'\tilde{\mathbf{D}}\right)^{-1} \tilde{\mathbf{D}}'\tilde{u} \tag{32}$$

Define $\mathbf{V} = \tilde{\mathbf{D}}'\tilde{\mathbf{D}}$ and $\mathbf{Z} = \tilde{\mathbf{D}}'\tilde{u}$. In calculating the bias, we will focus on results for large samples, where we deploy the weak law of large numbers and the continuous mapping theorem.

A.1.4 Computing V

Define the elements of the matrix $\mathbf{V}$ as

$$\mathbf{V} = \begin{pmatrix} V_{CC} & V_{CH} \\ V_{CH} & V_{HH} \end{pmatrix} \tag{33}$$

The term V_{CC} relates to the variance of $\tilde{C}_{it}$.

$$\begin{aligned} V_{CC} &= \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{C}_{it}^2 \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} (\tilde{\delta}_{C,i}\tilde{t} + \tilde{v}_{C,it})^2 \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\delta}_{C,i}^2\tilde{t}^2 + \sum_{i=1}^{N}\sum_{t=1}^{T} 2\tilde{\delta}_{C,i}\tilde{t}\tilde{v}_{C,it} \; + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{v}_{C,it}^2 \end{aligned}$$

Take the probability limits for large N, so that each summation over i converges to its expectation. Define $S_t = \sum_{t=1}^{T} \tilde{t}^2$ where time is deterministic. Note that $\mathbb{E}[\tilde{\delta}_{C,i}^2] = \text{Var}(\delta_{C,i})$, and $\mathbb{E}\left[\tilde{v}_{C,it}^2\right] = (1 - 1/T)\sigma_C^2$.[28] The (true) error terms, $v_{C,it}$ and $v_{H,it}$, are *iid*, so that $\mathbb{E}\left[\tilde{\delta}_{C,i}\tilde{v}_{C,it}\right] = 0$. Then it follows that

$$V_{CC} \to NS_t\text{Var}(\delta_{C,i}) + N(T-1)\sigma_C^2 \tag{34}$$

The term V_{HH} relates to the variance of $\tilde{H}_{it}$. Following similar steps we have

$$V_{HH} \to NS_t\text{Var}(\delta_{H,i}) + N(T-1)\sigma_H^2 \tag{35}$$

The term V_{CH} relates to the covariance between $\tilde{C}_{it}$ and $\tilde{H}_{it}$, which is computed as:

$$V_{CH} = \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{C}_{it}\tilde{H}_{it} \tag{36}$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} (\tilde{\delta}_{C,i}\tilde{t} + \tilde{v}_{C,it})(\tilde{\delta}_{H,i}\tilde{t} + \tilde{v}_{H,it}) \tag{37}$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\delta}_{C,i}\tilde{\delta}_{H,i}\tilde{t}^2 + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{t}\tilde{\delta}_{C,i}\tilde{v}_{H,it} + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{t}\tilde{\delta}_{H,i}\tilde{v}_{C,it} + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{v}_{C,it}\tilde{v}_{H,it} \tag{38}$$

Again, taking probability limits, and noting that the true errors (v) are *iid* and mean zero conditional on regressors, it follows that all summations converge to zero except the first, which is the covariance of $\tilde{\delta}_{C,i}$ and $\tilde{\delta}_{H,i}$. Thus

$$V_{CH} \to NS_t\text{Cov}(\delta_{C,i}, \delta_{H,i}) \tag{39}$$

[28] We are not (yet) assuming large T, so the empirical variance of the idiosyncratic disturbances that depend on both i and t are not identical to the true variance in short T panels.

A.1.5 Inverting the Matrix V

The inverse of $\mathbf{V}$ is:

$$\mathbf{V}^{-1} = \frac{1}{\det(\mathbf{V})} \begin{pmatrix} V_{HH} & -V_{CH} \\ -V_{CH} & V_{CC} \end{pmatrix} \tag{40}$$

where the determinant of $\mathbf{V}$ is:

$$\det(\mathbf{V}) = V_{CC}V_{HH} - V_{CH}^2 \tag{41}$$

Taking probability limits and using the above results, we have

$$\det(\mathbf{V}) \to \left(NS_t\text{Var}(\delta_{C,i}) + N(T-1)\sigma_C^2\right)\left(NS_t\text{Var}(\delta_{H,i}) + N(T-1)\sigma_H^2\right) - \left(NS_t\text{Cov}(\delta_{C,i}, \delta_{H,i})\right)^2 \tag{42}$$

$$= N^2S_t^2\left(\left(\text{Var}(\delta_{C,i}) + \frac{(T-1)\sigma_C^2}{S_t}\right)\left(\text{Var}(\delta_{H,i}) + \frac{(T-1)\sigma_H^2}{S_t}\right) - \left(\text{Cov}(\delta_{C,i}, \delta_{H,i})\right)^2\right) \tag{43}$$

A.1.6 Computing Z

The covariance vector $Z = \tilde{\mathbf{D}}'\tilde{u}$ has elements

$$\mathbf{Z} = \begin{pmatrix} Z_C \\ Z_H \end{pmatrix} \tag{44}$$

which we compute as follows.

$$Z_C = \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{C}_{it}\tilde{u}_{it} \tag{45}$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} (\tilde{\delta}_{C,i}\tilde{t} + \tilde{v}_{C,it})(\tilde{\gamma}_i\tilde{t} + \tilde{\varepsilon}_{it}) \tag{46}$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{t}^2\tilde{\delta}_{C,i}\tilde{\gamma}_i + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\delta}_{C,i}\tilde{\varepsilon}_{it}\tilde{t} + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\gamma}_i\tilde{v}_{C,it}\tilde{t} + \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{v}_{C,it}\tilde{\varepsilon}_{it} \tag{47}$$

Taking probability limits, and noting that the true errors (v and ϵ) are *iid* and mean zero conditional on regressors, it follows that all summations converge to zero except the first, which incorporates the covariance of $\tilde{\delta}_{C,i}$ and $\tilde{\gamma}_i$. Thus

$$Z_C \to NS_t \text{Cov}(\delta_{C,i}, \gamma_i) \tag{48}$$

Similarly for Z_H.

$$Z_H \to NS_t \text{Cov}(\delta_{H,i}, \gamma_i) \tag{49}$$

A.1.7 Computing Bias

Recall that the bias is:

$$\begin{pmatrix} \text{Bias}(\hat{\beta}_C) \\ \text{Bias}(\hat{\beta}_H) \end{pmatrix} = \mathbf{V}^{-1}\mathbf{Z} \tag{50}$$

Using the matrix definitions, we have

$$\text{Bias}(\hat{\beta}_C) = \frac{1}{\det(\mathbf{V})} \left(V_{HH} Z_C - V_{CH} Z_H \right) \tag{51}$$

Using the probability limits calculated for each term above, and assuming a non-zero determinant (full rank), this bias converges to

$$\text{Bias}(\hat{\beta}_C) \to \frac{\left(NS_t \text{Var}(\delta_{H,i}) + N(T-1)\sigma_H^2\right)\left(NS_t \text{Cov}(\delta_{C,i}, \gamma_i)\right) - \left(NS_t \text{Cov}(\delta_{C,i}, \delta_{H,i})\right)\left(NS_t \text{Cov}(\delta_{H,i}, \gamma_i)\right)}{N^2 S_t^2 \left(\left(\text{Var}(\delta_{C,i}) + \frac{(T-1)\sigma_C^2}{S_t} \right) \left(\text{Var}(\delta_{H,i}) + \frac{(T-1)\sigma_H^2}{S_t} \right) - \left(\text{Cov}(\delta_{C,i}, \delta_{H,i}) \right)^2 \right)}$$

Or, simplifying, we can write this in terms of variances and covariances as

$$\text{Bias}(\hat{\beta}_C) \to \frac{\left(\text{Var}(\delta_{H,i}) + \frac{T-1}{S_t}\sigma_H^2 \right) \left(\text{Cov}(\delta_{C,i}, \gamma_i) \right) - \left(\text{Cov}(\delta_{C,i}, \delta_{H,i}) \right) \left(\text{Cov}(\delta_{H,i}, \gamma_i) \right)}{\left(\text{Var}(\tilde{\delta}_{C,i}) + \frac{T-1}{S_t}\sigma_C^2 \right) \left(\text{Var}(\delta_{H,i}) + \frac{T-1}{S_t}\sigma_H^2 \right) - \left(\text{Cov}(\delta_{C,i}, \delta_{H,i}) \right)^2} \tag{52}$$

Similarly, we calculate $\text{Bias}(\hat{\beta}_H)$ as

$$\text{Bias}(\hat{\beta}_H) = \frac{1}{\det(\mathbf{V})} \left(V_{CC} Z_H - V_{CH} Z_C \right) \tag{53}$$

which converges for large N as

$$\text{Bias}(\hat{\beta}_H) \to \frac{\left(\text{Var}(\delta_{C,i}) + \frac{T-1}{S_t}\sigma_C^2\right)(\text{Cov}(\delta_{H,i},\gamma_i)) - (\text{Cov}(\delta_{C,i},\delta_{H,i}))(\text{Cov}(\delta_{C,i},\gamma_i))}{\left(\text{Var}(\delta_{C,i}) + \frac{T-1}{S_t}\sigma_C^2\right)\left(\text{Var}(\delta_{H,i}) + \frac{T-1}{S_t}\sigma_H^2\right) - (\text{Cov}(\delta_{C,i},\delta_{H,i}))^2} \tag{54}$$

A.1.8 Bias with Large T

For large T, S_t grows on the order of T^3. Therefore, the terms involving T/S_t become negligible compared to other terms. Simplifying on these grounds, we then have

$$\text{Bias}(\hat{\beta}_C) = \frac{\text{Var}(\delta_{H,i})\text{Cov}(\delta_{C,i},\gamma_i) - \text{Cov}(\delta_{C,i},\delta_{H,i})\text{Cov}(\delta_{H,i},\gamma_i)}{\text{Var}(\delta_{C,i})\text{Var}(\delta_{H,i}) - (\text{Cov}(\delta_{C,i},\delta_{H,i}))^2} \tag{7}$$

$$\text{Bias}(\hat{\beta}_H) = \frac{\text{Var}(\delta_{C,i})\text{Cov}(\delta_{H,i},\gamma_i) - \text{Cov}(\delta_{C,i},\delta_{H,i})\text{Cov}(\delta_{C,i},\gamma_i)}{\text{Var}(\delta_{C,i})\text{Var}(\delta_{H,i}) - (\text{Cov}(\delta_{C,i},\delta_{H,i}))^2} \tag{8}$$

as was to be shown.

A.1.9 Bias Map

With these results, we can consider the direction of the bias. Rearranging the numerators of the bias equations (7) and (8), define the ratios

$$R_A = \frac{\text{Var}(\delta_{H,i})\text{Cov}(\delta_{C,i},\gamma_i)}{\text{Cov}(\delta_{C,i},\delta_{H,i})\text{Cov}(\delta_{H,i},\gamma_i)} \tag{55}$$

$$R_B = \frac{\text{Var}(\delta_{C,i})\text{Cov}(\delta_{H,i},\gamma_i)}{\text{Cov}(\delta_{C,i},\delta_{H,i})\text{Cov}(\delta_{C,i},\gamma_i)} \tag{56}$$

Further note that that a positive determinant of the matrix $\boldsymbol{V}$ implies that

$$R_A R_B > 1 \tag{57}$$

The potential bias cases are summarized in Figure 2. ■

A.2 Proof of Result 2

Consider the trend processes given by (9), (10), and (11), as repeated here:

$$\begin{aligned}\delta_{C,i} &= \psi_C + \omega_C T_{0,i} + e_{C,i} \\ \delta_{H,i} &= \psi_H + \omega_H T_{0,i} + e_{H,i} \\ \gamma_i &= \psi_Y + \omega_Y T_{0,i} + e_{Y,i}\end{aligned}$$

It then follows that the covariances are

$$\begin{aligned}\text{Cov}(\delta_{C,i}, \delta_{H,i}) &= \omega_C \omega_H \sigma^2_{T_0} \\ \text{Cov}(\delta_{C,i}, \gamma_i) &= \omega_C \omega_Y \sigma^2_{T_0} \\ \text{Cov}(\delta_{H,i}, \gamma_i) &= \omega_H \omega_Y \sigma^2_{T_0}\end{aligned}$$

and the variances are

$$\begin{aligned}\text{Var}(\delta_{C,i}) &= \omega^2_C \sigma^2_{T_0} + \sigma^2_{e_C} \\ \text{Var}(\delta_{H,i}) &= \omega^2_H \sigma^2_{T_0} + \sigma^2_{e_H}\end{aligned}$$

Plug these variance and covariance results in to the finite T results above (equations (52) and (54)). Collecting terms and simplifying leads to the forms:

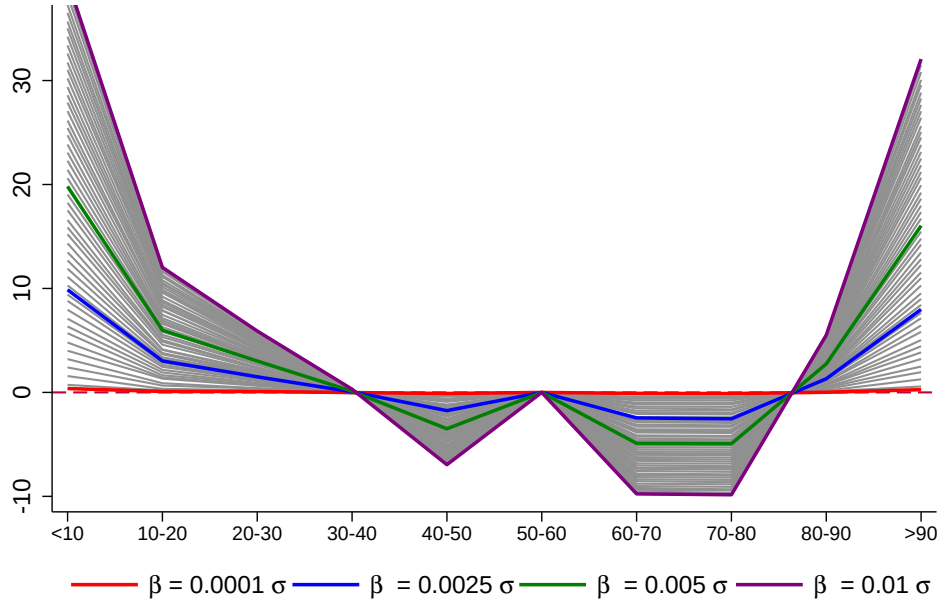
$$\text{Bias}(\hat{\beta}_C) = \frac{\omega_C \omega_Y}{\omega^2_C + \frac{A_C}{A_H}\omega^2_H + A_H} \tag{58}$$

$$\text{Bias}(\hat{\beta}_H) = \frac{\omega_H \omega_Y}{\omega^2_H + \frac{A_H}{A_C}\omega^2_H + A_C} \tag{59}$$

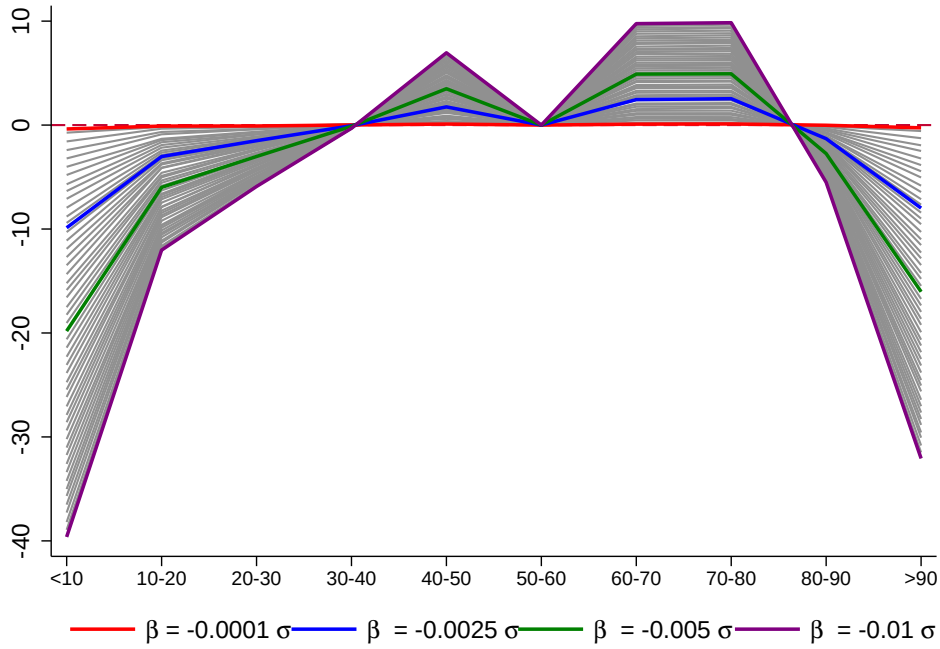
where we define $A_C = \frac{\sigma^2_{e_C}}{\sigma^2_{T_0}} + \frac{T-1}{S_t}\frac{\sigma^2_C}{\sigma^2_{T_0}}$ and $A_H = \frac{\sigma^2_{e_H}}{\sigma^2_{T_0}} + \frac{T-1}{S_t}\frac{\sigma^2_H}{\sigma^2_{T_0}}$.

Since $\omega_C > 0$ and $\omega_H > 0$ and all terms in the denominator are positive, it follows by inspection that the biases are both positive (a U-shape) if $\omega_Y > 0$ and the biases are both negative (an inverted U-shape) if $\omega_Y < 0$. Further, by inspection, there is no bias if $\omega_Y = 0$.

■



(a) Spurious Effects by Magnitude of Positive Outcome Trend



(b) Spurious Effects by Magnitude of Negative Outcome Trend

Figure A.1: Magnitudes of Outcome Trend and Spurious Effects

Notes: Each sub-figure reports estimates from equation (1), in which the outcome is simulated using the following equation: $Y_{ct} = \beta \cdot t \cdot Temp_c^{base} + \epsilon_{ct}$, $\epsilon_{ct} \sim \mathcal{N}(0, \sigma)$, and σ equals the standard deviation of $tTemp_c^{base}$ over the full sample. The temperature bins are constructed from historical, daily temperature realizations using the ERA-5 Land dataset. In Panel (a), the outcome is simulated to have a positive differential trend in baseline temperature (i.e., $\beta \in [0.0001\sigma, 0.01\sigma]$); in Panel (b), the outcome is simulated to have a negative differential trend in baseline temperature (i.e., $\beta \in [-0.01\sigma, -0.0001\sigma]$). In each sub-figure, we report the mean of each coefficient from (1) taking 1,000 separate draws of the outcome variable for each value of β and highlight four examples in red ($\beta = \pm 0.0001\sigma$), blue ($\beta = \pm 0.0025\sigma$), green ($\beta = \pm 0.005\sigma$) and purple ($\beta = \pm 0.01\sigma$).

.

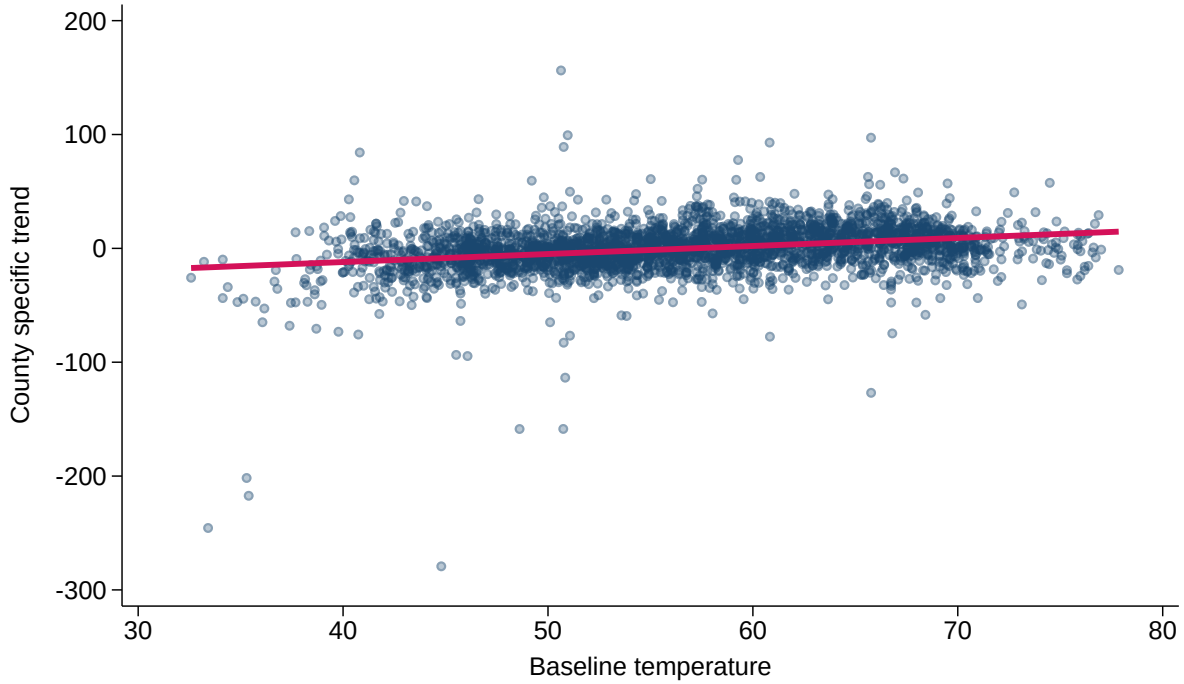
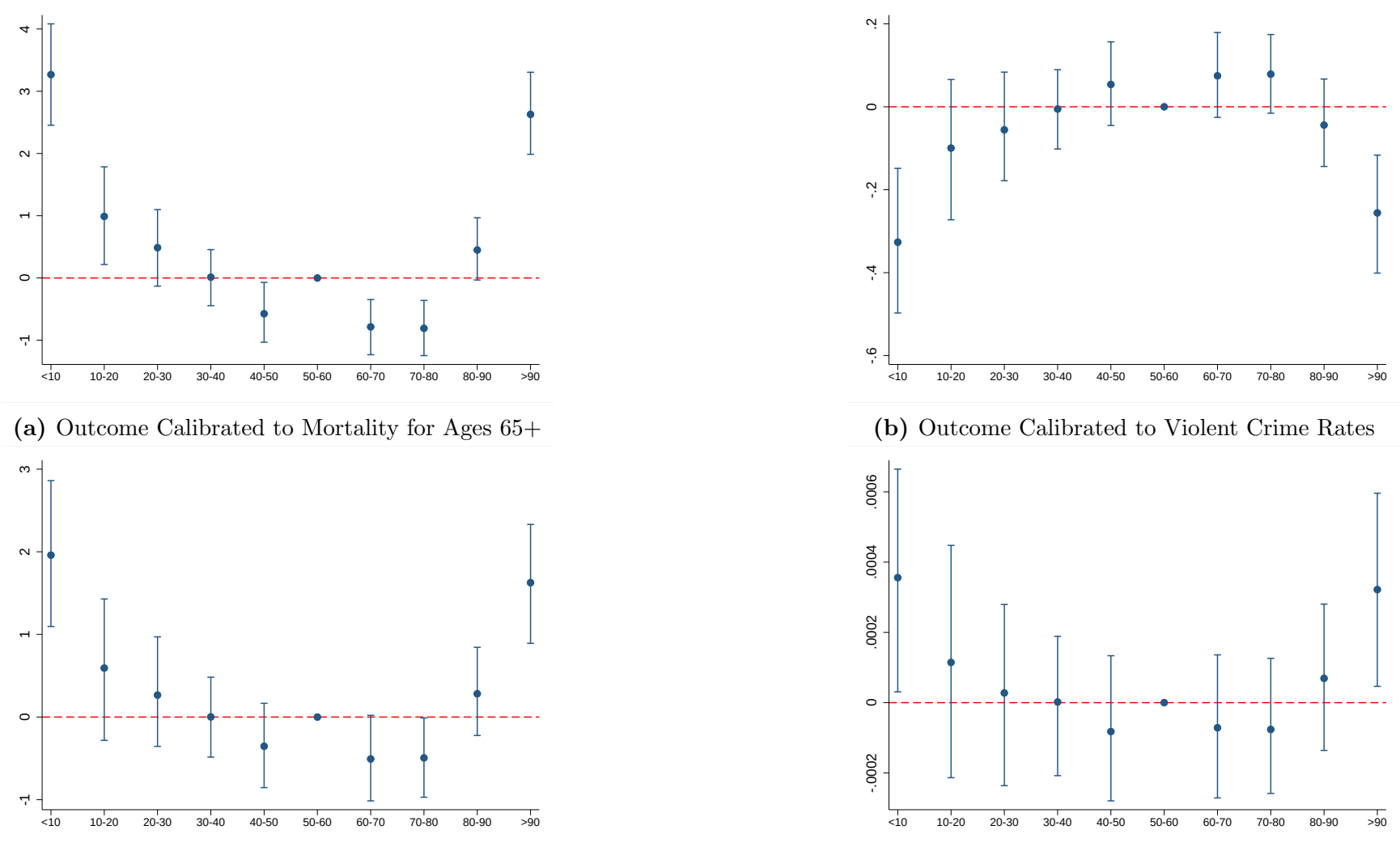


Figure A.2: Mortality Trend vs. Baseline Temperature

Notes: This figure reports the cross-county relationship between baseline temperature ($Temp_c^{Base}$) and county-specific trends ($\hat{\alpha_c}$) from equation (17), when the outcome variable is the county-level annual mortality rate for people age 65 and over (per 100,000 inhabitants, for the period 1970-2002). $Temp_c^{Base}$ is constructed from historical, daily temperature realizations using the ERA-5 Land dataset and corresponds to the average daily temperature in the 1970s. Mortality rate for individuals over the age of 65 is measured using the CDC morality files.



(a) Outcome Calibrated to Mortality for Ages 65+

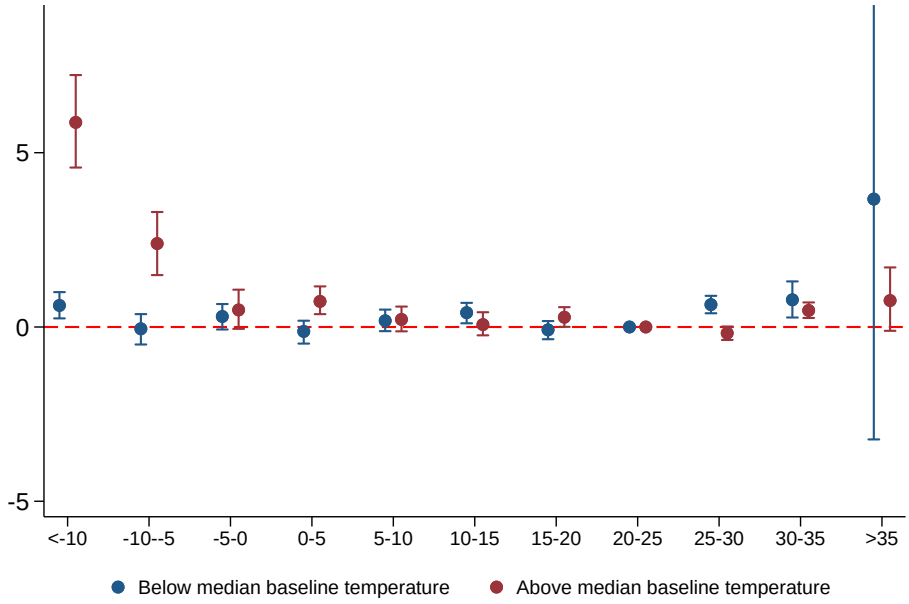
(b) Outcome Calibrated to Violent Crime Rates

(c) Outcome Calibrated to Nonviolent Crime Rates

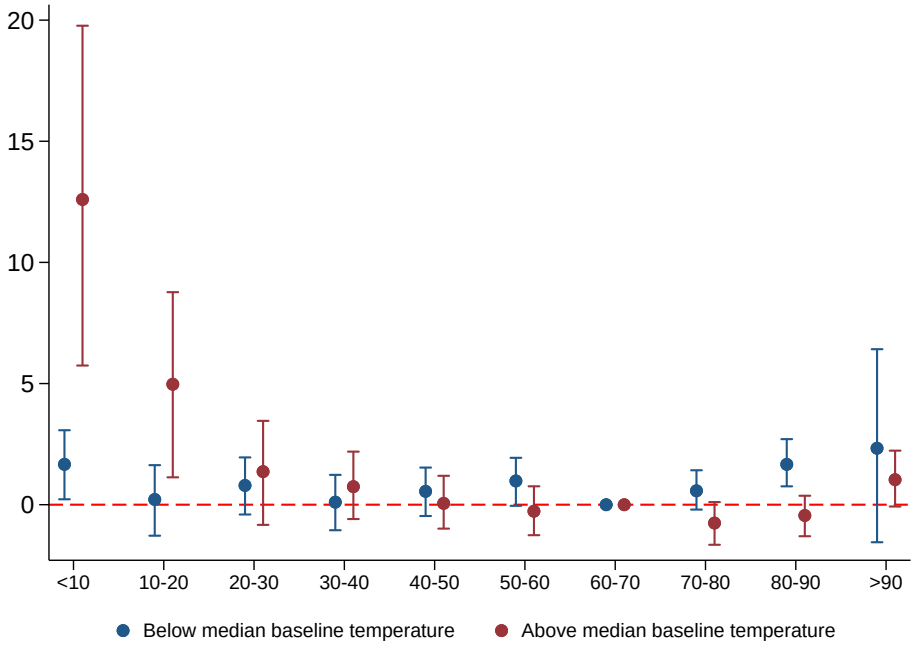
(d) Outcome Calibrated to Corn Yields

Figure A.3: Simulation with Calibrated Outcomes and Real Temperature

Notes: Each sub-figure reports estimates from equation (1), in which the outcome is calibrated to real data using equation (19). The temperature bins are constructed from historical, daily temperature realizations using the ERA-5 Land dataset. In each sub-figure, we report the mean and 95% confidence intervals of each coefficient from (1) taking 1,000 separate draws of the outcome variable. In Panel (a), the outcome is calibrated to the mortality rate for people age 65 and over using CDC mortality files. In Panels (b) and (c), the outcome is calibrated to violent/non violent crime rate respectively (i.e., murder, rape, and aggravated assault for Panel (b) and larceny and robbery for Panel (c)) measured using data from the FBI. In Panel (d), the outcome variable is calibrated to log of corn yield, measured using NASS Quickstats.



(a) Binned Regression Estimates: Celsius

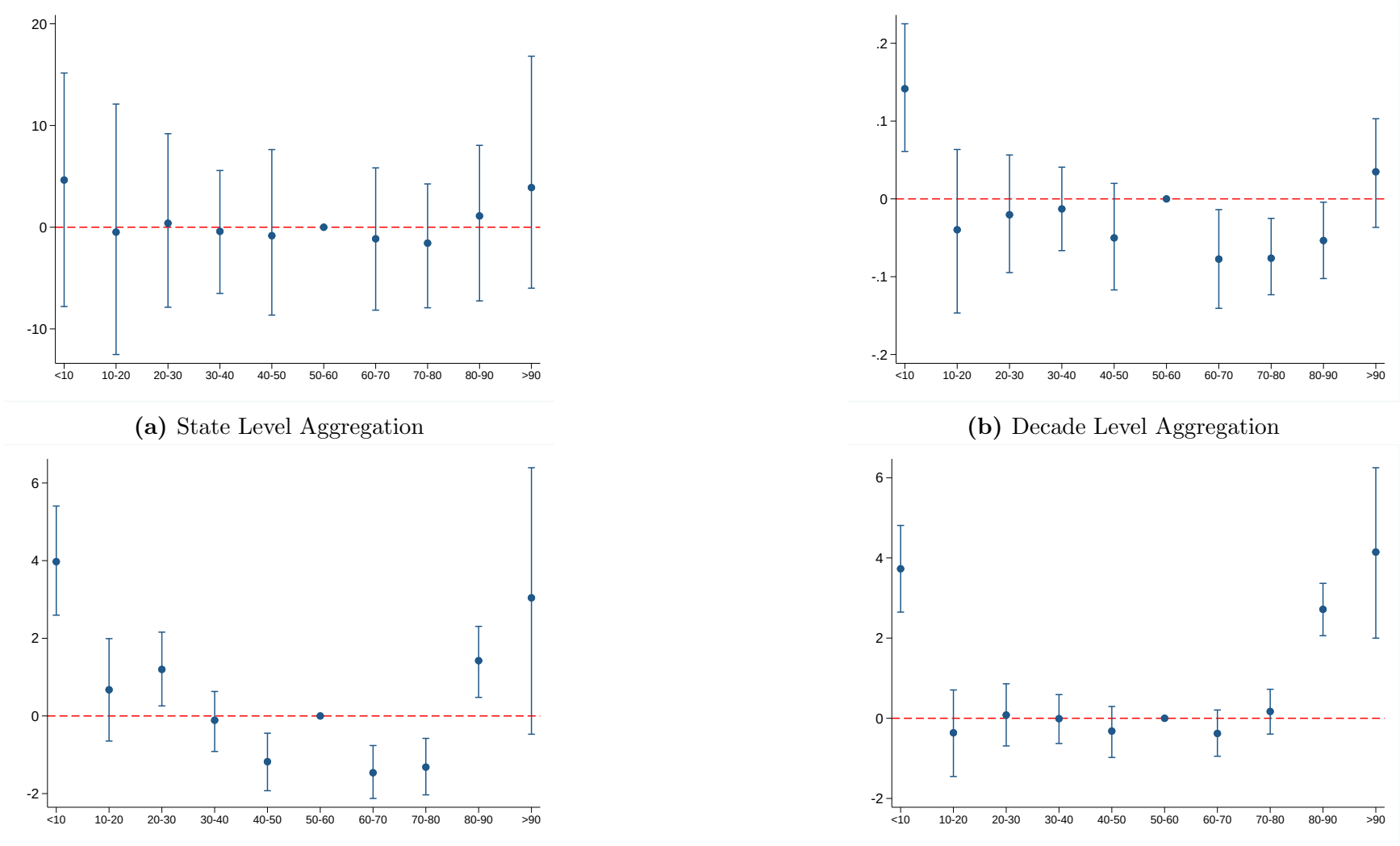


(b) Binned Regression Estimates: Omitting the 60-70°F Bin

Figure A.4: Binned Regression Estimates and "Adaptation": Alternative Binning

Notes: We report estimates of equation (1), separately for counties with above versus below median baseline temperature. The outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. In Panel (a), the dataset is constructed in Celsius, and the 20-25°C (68-77°F) bin is omitted for normalization. In Panel (b), the dataset is in Fahrenheit but we omit the 60-70°F bin instead of the 50-60°F bin as in Figure 6c. We report the mean and 95% confidence intervals for each coefficient estimate, taking 1,000 draws of ϵ_{ct} from a normal distribution centered at zero with variance equivalent to the in-sample variance of $tTemp_c^{base}$.

.



(a) State Level Aggregation

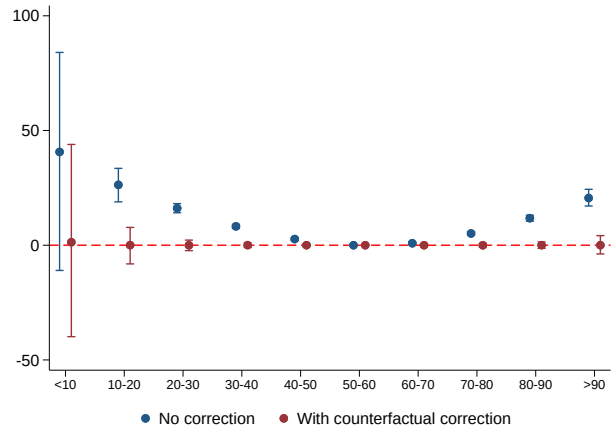
(b) Decade Level Aggregation

(c) PRISM Temperature Data

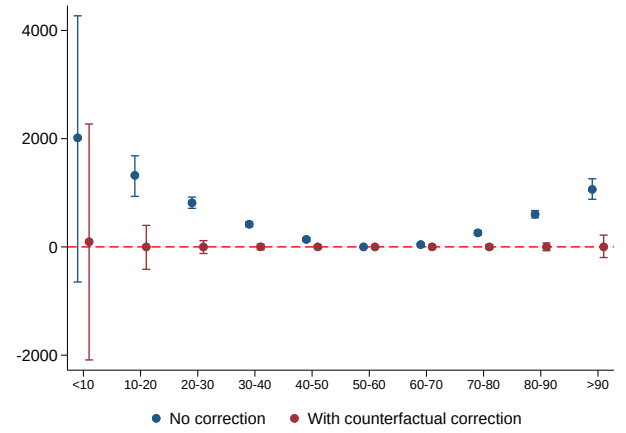
(d) GHCN Temperature Data

Figure A.5: Binning Bias: Sensitivity and Extensions to Alternative Levels of Aggregation and Temperature Datasets

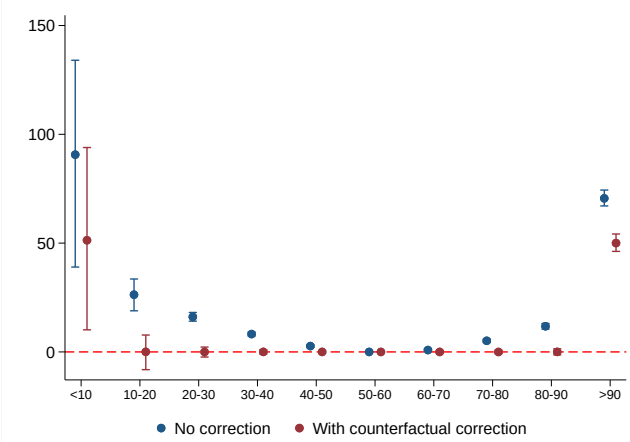
Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data, under a variety of aggregation levels and temperature datasets. Each sub-figure reports estimates of equation (1). The outcome is simulated to have a positive differential trend in baseline temperature ($Y_{ct} = tTemp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset for Panels (a) and (b). We report the mean and 95% confidence intervals for each coefficient estimate, taking 1,000 draws of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to the in-sample variance of $tTemp_c^{base}$. Panel (a) presents results when the outcome and temperature variables are aggregated at the state level; Panel (b) presents results when the outcome and temperature variables are aggregated at the decade level. Panels (c) and (d) present results when daily temperature realizations are from PRISM and GHCN datasets, respectively.



(a) Counterfactual Correction: Linear Trend



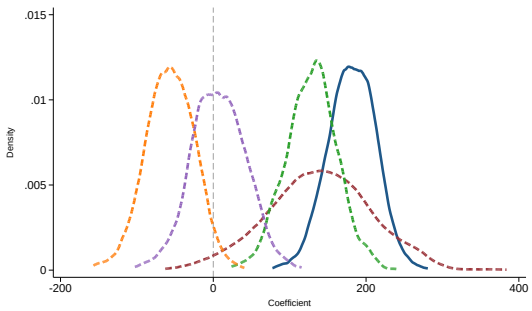
(b) Counterfactual Correction: Quadratic Trend



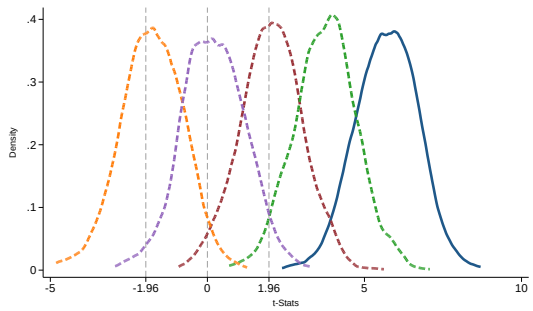
(c) Counterfactual Correction: True Effects

Figure A.6: Counterfactual Temperatures as a Solution to Binning Bias

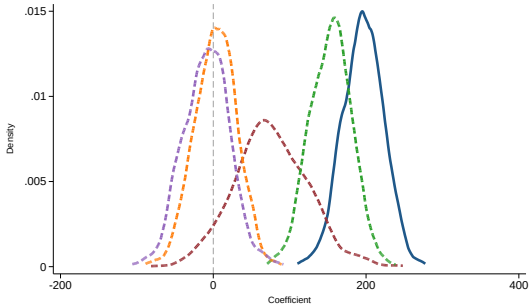
Notes: Each sub-figure reports estimates of equation (1), alongside of estimates of the same equation that include the expected counterfactual exposure to each temperature bin as controls. In Panels (a) and (c), the outcome is simulated to have a differential linear trend in baseline temperature, $Y_{ct} = t \cdot Temp_c^{base} + \epsilon_{ct}$. The error term is drawn from a normal distribution centered at zero with variance equal to the variance of $tTemp_c^{base}$ over the full sample. In Panel (b), the outcome is simulated to have a differential quadratic trend in baseline temperature $Y_{ct} = t^2 \cdot T_c^{Base} + \epsilon_{ct}$. The error term is drawn from a normal distribution centered at zero with variance equal to the variance of $t^2 Temp_c^{base}$ over the full sample. The temperature data are drawn so that all counties have the same warming trend (1°C over the sample period), starting with a normal distribution centered at the county-level baseline temperature. In Panels (a) and (c) the 1°C increase follows a linear trend over the period, while in Panel (b) the 1°C increase follows a quadratic trend. In Panel (c), the outcome has the same differential trend in the outcome variable as Panel (a) but there is also a positive effect of 50 of exposure to the top and bottom temperature bins. We report the mean and 95% confidence intervals for each coefficient estimate from (1), taking 1,000 draws of the outcome variable.



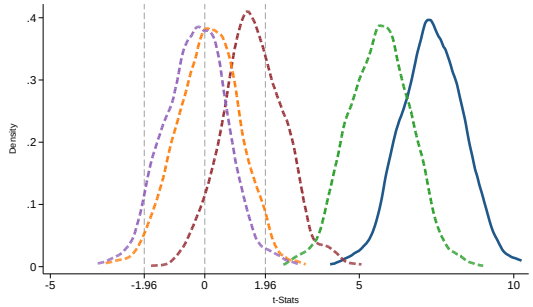
(a) Under 10 °F Bin: Coefficient



(b) Under 10 °F Bin: t-Stat



(c) Over 90 °F Bin: Coefficient



(d) Over 90 °F Bin: t-Stat

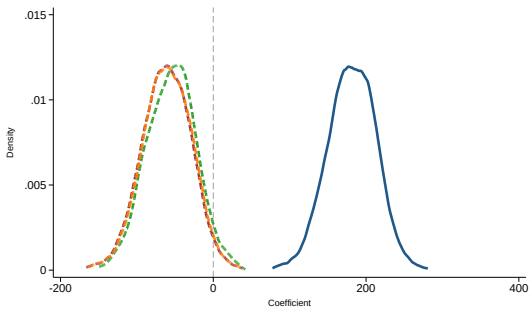
Figure A.7: Comparison of Reduced Form Solutions with Simulated Outcome Variable (Quadratic Trends)

—— No correction (Under 10 Bin: 100.0%, Over 90 Bin: 100.0%) - - - State-Year Fixed Effects (31.2%, 71.5%)

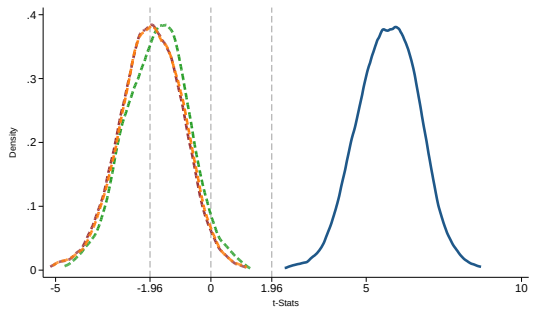
- - - With 3 Lags (100.0%, 100.0%) - - - County-Specific Linear Trends (40.5%, 4.5%)

- - - County-5 Year Fixed Effects (14.3%, 5.6%)

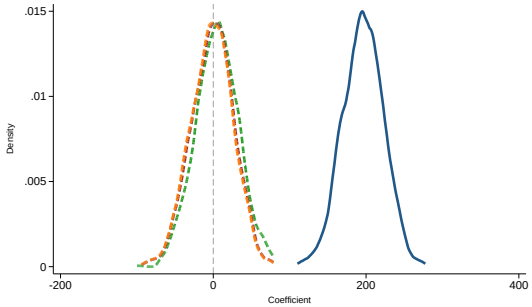
Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data. Here, we plot the density of coefficients and t-statistics from estimating equation (5) 1,000 times, which only uses the two extreme temperature bins, alongside densities from specifications that use alternative reduced-form corrective strategies as indicated in the legend. The outcome is simulated to have a positive quadratic differential trend in baseline temperature ($Y_{ct} = t^2 Temp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. Each time, we take a draw of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to two times the in-sample standard deviation of $t^2 Temp_c^{base}$. Numbers inside the parentheses in the legend denote the number of runs per 1,000 simulations that the coefficient is significant at the 5% level, for the under 10°F bin and the over 90°F bin respectively.



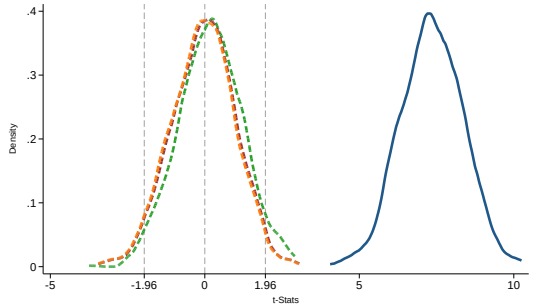
(a) Under 10 Bin: Coefficient



(b) Under 10 Bin: t-Stat



(c) Over 90 Bin: Coefficient



(d) Over 90 Bin: t-Stat

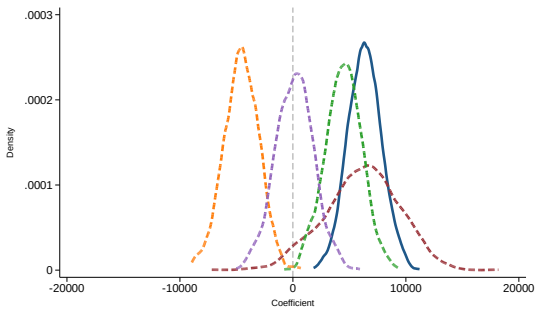
Figure A.8: Comparison of `cftemp` Solutions with Simulated Outcome Variable (Quadratic Trends)

—— No correction (Under 10 Bin: 100.0%, Over 90 Bin: 100.0%)

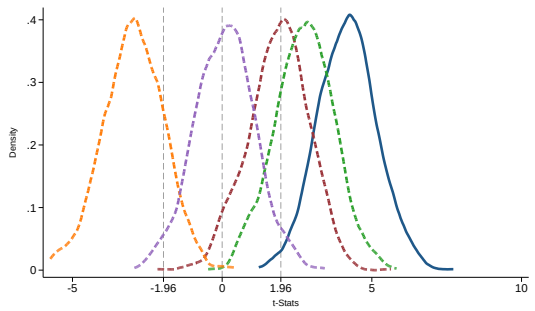
- - - Counterfactual Temperature Control (33.6%, 5.2%)

- - - Counterfactual Temperature Control without Bayes (19.3%, 5.9%) - - - Chebyshev (22.0%, 5.2%)

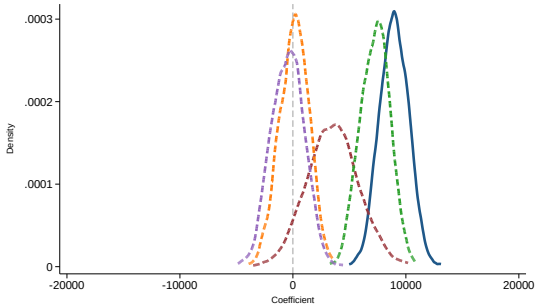
Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data. Here, we plot the density of coefficients and t-statistics from estimating equation (5) 1,000 times, which only uses the two extreme temperature bins, alongside densities from specifications that also include the expected counterfactual exposure to each temperature bin as controls. Each color and line plots densities that use different methods to construct the counterfactuals, as indicated in the legend. The outcome is simulated to have a positive quadratic differential trend in baseline temperature ($Y_{ct} = t^2 Temp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. Each time, we take a draw of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to two times the in-sample standard deviation of $t^2 Temp_c^{base}$. Numbers inside the parentheses in the legend denote the number of runs per 1,000 simulations that the coefficient is significant at the 5% level, for the under 10°F bin and the over 90°F bin respectively.



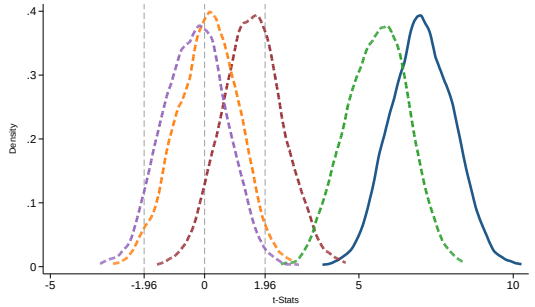
(a) Under 10 °F Bin: Coefficient



(b) Under 10 °F Bin: t-Stat



(c) Over 90 °F Bin: Coefficient



(d) Over 90 °F Bin: t-Stat

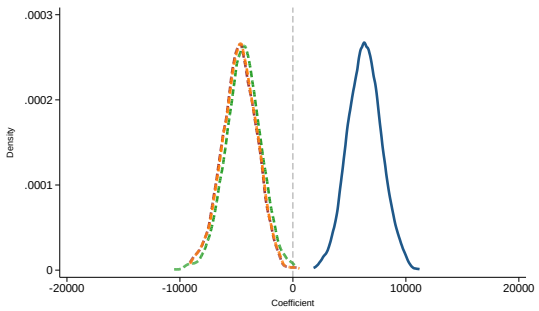
Figure A.9: Comparison of Reduced Form Solutions with Simulated Outcome Variable (Cubic Trends)

—— No correction (Under 10 Bin: 100.0%, Over 90 Bin: 100.0%) - - - State-Year Fixed Effects (21.3%, 68.3%)

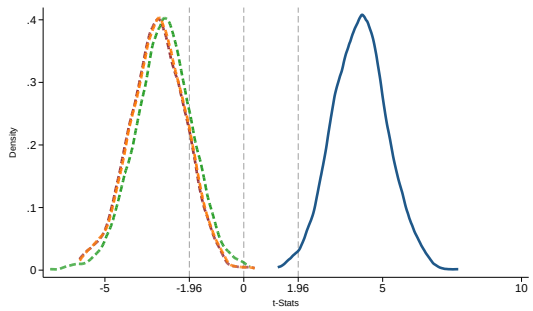
- - - With 3 Lags (99.0%, 100.0%) - - - County-Specific Linear Trends (81.2%, 6.0%)

- - - County-5 Year Fixed Effects (13.5%, 5.4%)

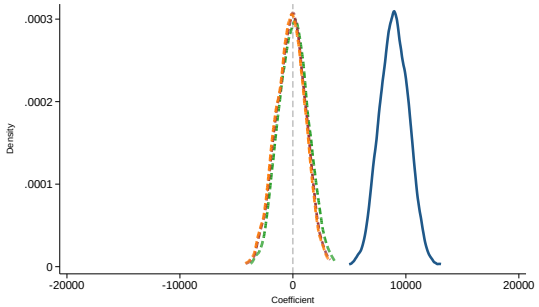
Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data. Here, we plot the density of coefficients and t-statistics from estimating equation (5) 1,000 times, which only uses the two extreme temperature bins, alongside densities from specifications that use alternative reduced-form corrective strategies as indicated in the legend. The outcome is simulated to have a positive cubic differential trend in baseline temperature ($Y_{ct} = t^3 Temp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. Each time, we take a draw of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to two times the in-sample standard deviation of $t^3 Temp_c^{base}$. Numbers inside the parentheses in the legend denote the number of runs per 1,000 simulations that the coefficient is significant at the 5% level, for the under 10°F bin and the over 90°F bin respectively.



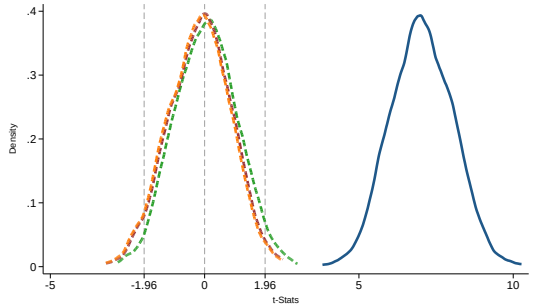
(a) Under 10 °F Bin: Coefficient



(b) Under 10 °F Bin: t-Stat



(c) Over 90 °F Bin: Coefficient



(d) Over 90 °F Bin: t-Stat

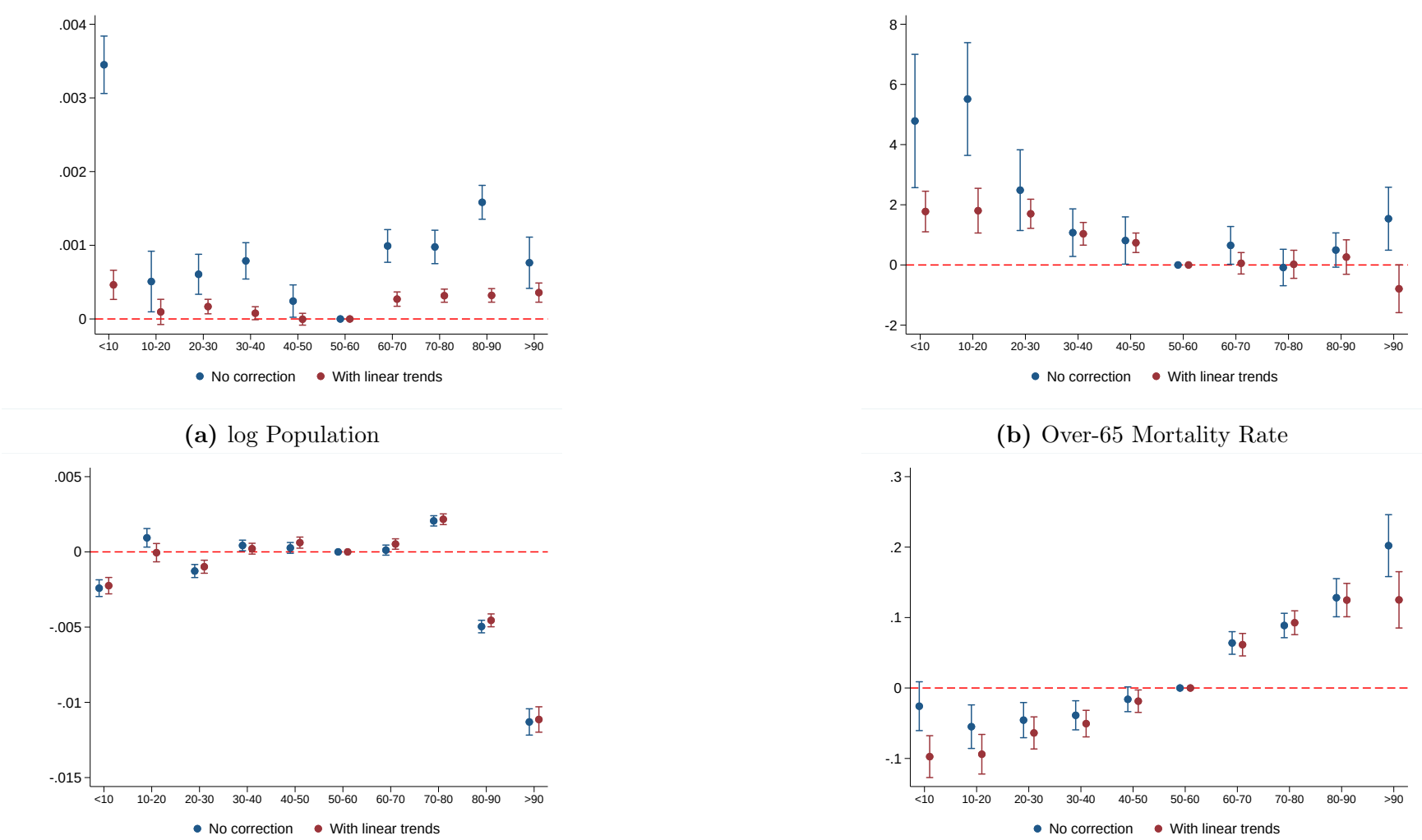
Figure A.10: Comparison of `cftemp` Solutions with Simulated Outcome Variable (Cubic Trends)

—— No correction (Under 10 Bin: 100.0%, Over 90 Bin: 100.0%)

- - - Counterfactual Temperature Control (76.2%, 6.8%)

- - - Counterfactual Temperature Control without Bayes (61.3%, 6.2%) - - - Chebyshev (64.7%, 7.9%)

Notes: We conduct the same exercise as Figure 5, which uses simulated outcome data and real temperature data. Here, we plot the density of coefficients and t-statistics from estimating equation (5) 1,000 times, which only uses the two extreme temperature bins, alongside densities from specifications that also include the expected counterfactual exposure to each temperature bin as controls. Each color and line plots densities that use different methods to construct the counterfactuals, as indicated in the legend. The outcome is simulated to have a positive cubic differential trend in baseline temperature ($Y_{ct} = t^3 Temp_c^{base} + \epsilon_{ct}$) and daily temperature realizations are from the ERA-5 Land dataset. Each time, we take a draw of $\epsilon_{c,t}$ from a normal distribution centered at zero with variance equivalent to two times the in-sample standard deviation of $t^3 Temp_c^{base}$. Numbers inside the parentheses in the legend denote the number of runs per 1,000 simulations that the coefficient is significant at the 5% level, for the under 10°F bin and the over 90°F bin respectively.



(a) log Population

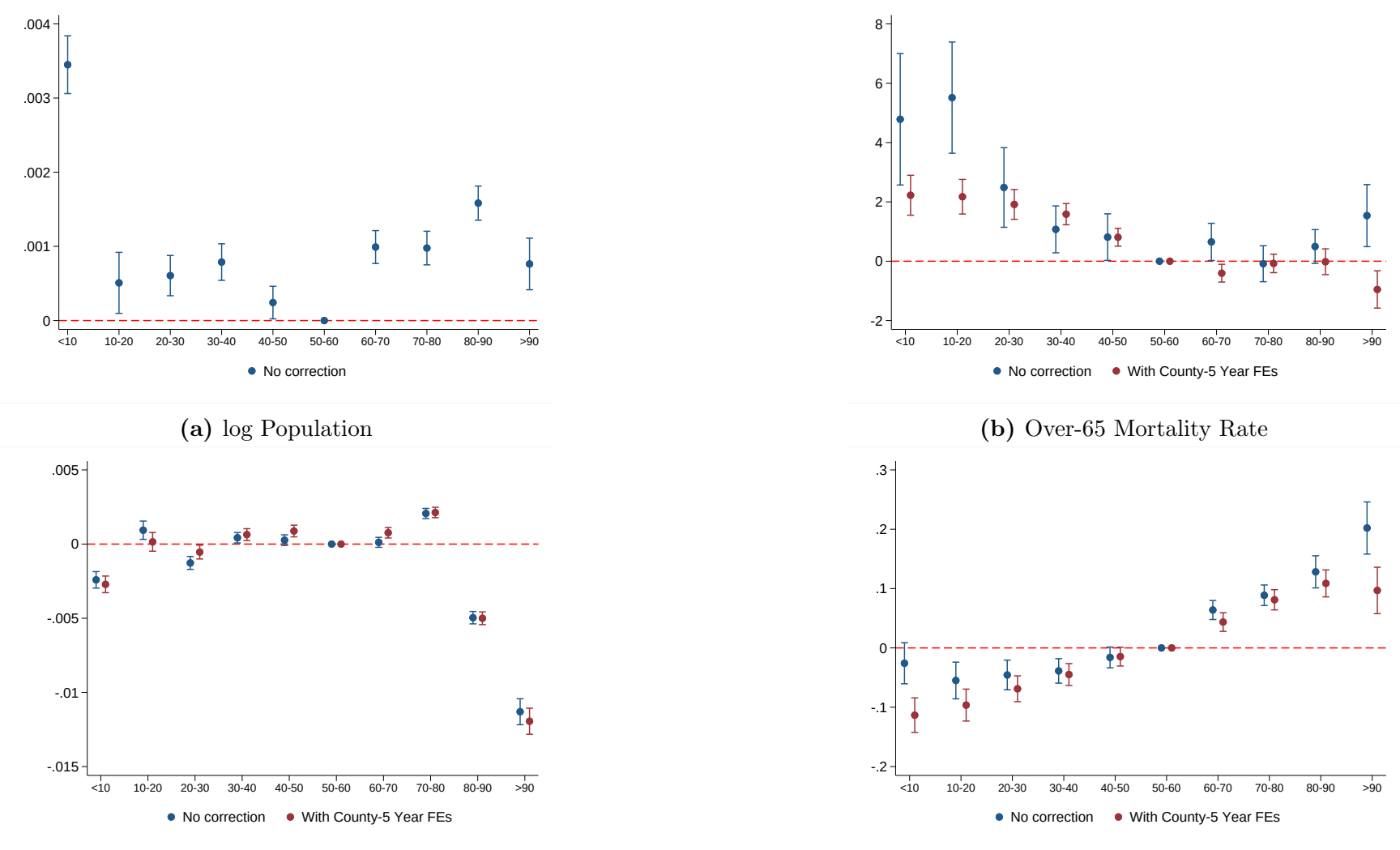
(b) Over-65 Mortality Rate

(c) log Corn Yield

(d) Violent Crime Rate

Figure A.11: Binning Bias with Real Outcome Data (County Specific Linear Trends)

Notes: This figure reports estimates of (1) for a series of outcome variables, alongside estimate of (21) that also include the county-specific linear trends as controls. Daily temperature realizations in each county used to construct each temperature bins are from the ERA-5 Land dataset. In Panel (a), the unit of observation is a county-decade and the outcome is log of population measured from the U.S. Census (1970-2010). In Panel (b), the unit of observation is a county-year (1970-2002) and the outcome variable is the mortality rate for individuals over the age of 65 (per 100,000 inhabitants), measured using the CDC mortality files. In Panel (c), the unit of observation is a county-year (1985-2006) and the outcome variable is log of corn yield, measured using NASS Quickstats. In Panel (d), the unit of observation is a county-month (1970-2009) and the outcome variable is violent crime rate (i.e., murder, rape, and aggravated assault) measured using data from the FBI. 95% confidence intervals are reported and standard errors are clustered by county.



(a) log Population

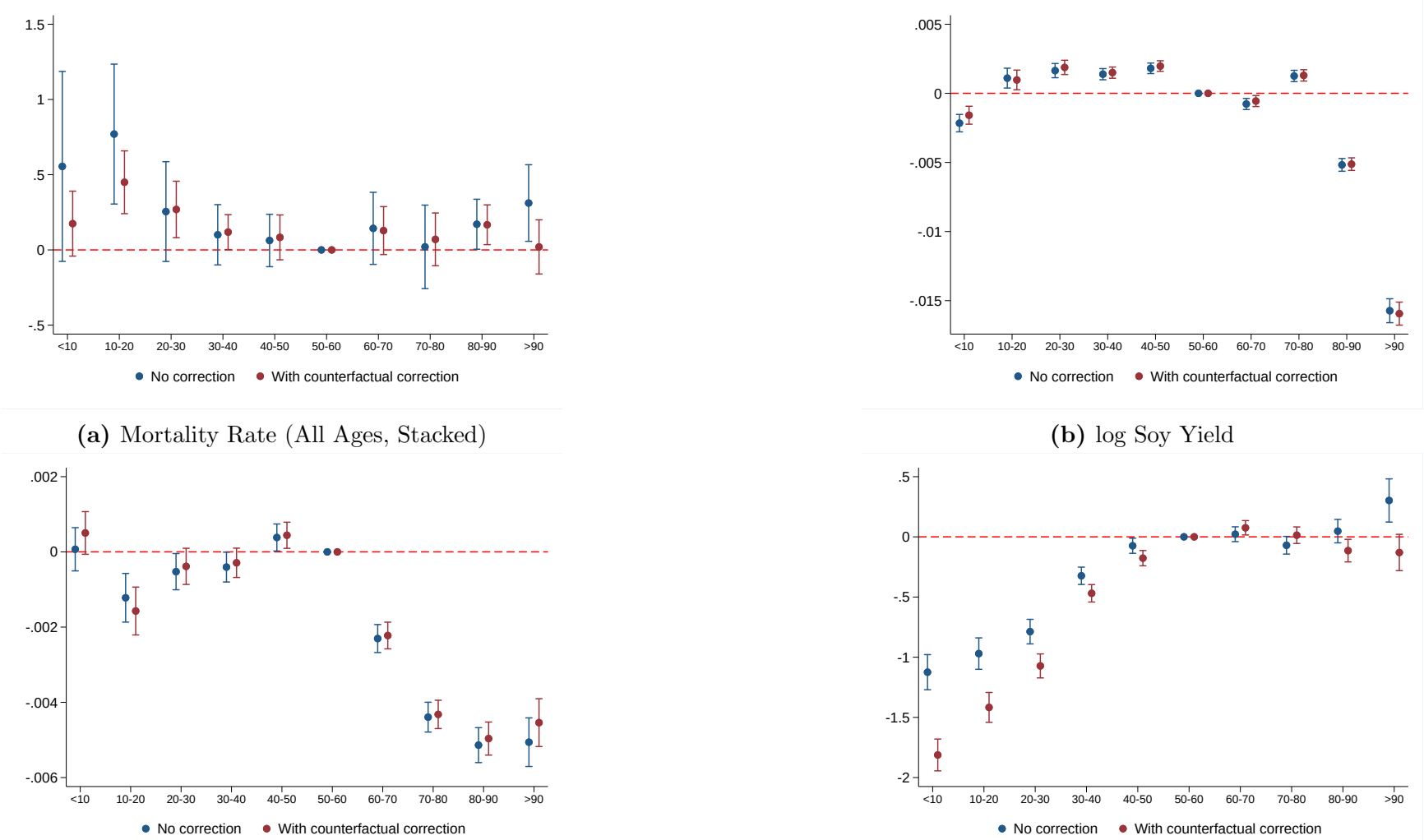
(b) Over-65 Mortality Rate

(c) log Corn Yield

(d) Violent Crime Rate

Figure A.12: Binning Bias with Real Outcome Data (County-5 Year Fixed Effects)

Notes: This figure reports estimates of (1) for a series of outcome variables, alongside estimates that also include county-5 year fixed effects as controls. Daily temperature realizations in each county used to construct each temperature bins are from the ERA-5 Land dataset. In Panel (a), the outcome is log of population measured from the U.S. Census (1970-2010). We do not report estimates with county-5-year fixed effects because the unit of observation is county-decade. In Panel (b), the unit of observation is a county-year (1970-2002) and the outcome variable is the mortality rate for individuals over the age of 65 (per 100,000 inhabitants), measured using the CDC mortality files. In Panel (c), the unit of observation is a county-year (1985-2006) and the outcome variable is log of corn yield, measured using NASS Quickstats. In Panel (d), the unit of observation is a county-month (1970-2009) and the outcome variable is violent crime rate (i.e., murder, rape, and aggravated assault) measured using data from the FBI. 95% confidence intervals are reported and standard errors are clustered by county.



(a) Mortality Rate (All Ages, Stacked)

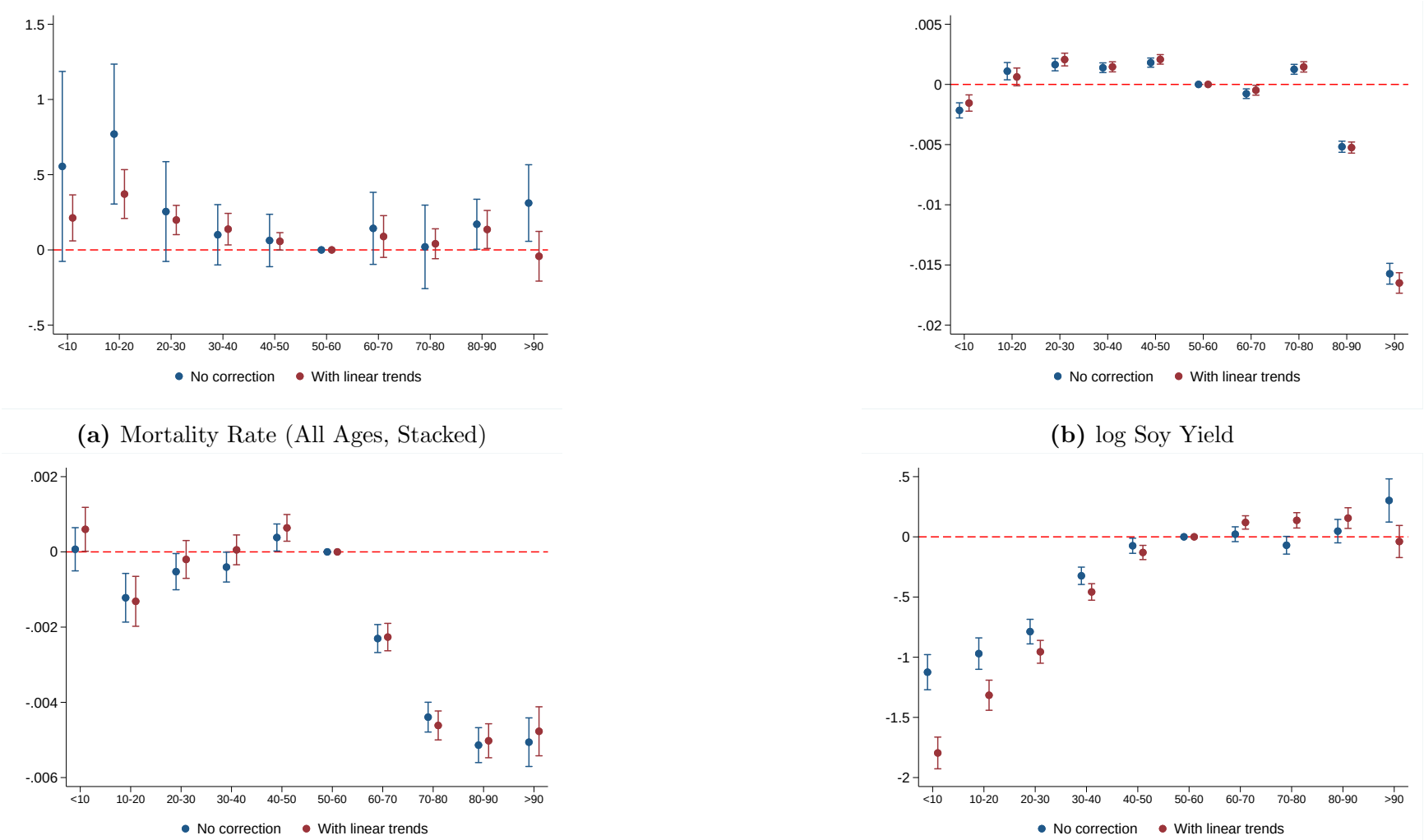
(b) log Soy Yield

(c) log Wheat Yield

(d) Nonviolent Crime

Figure A.13: Binning Bias with Real Outcome Data: Additional Outcomes (Counterfactual Controls with Empirical Bayes)

Notes: This figure reports estimates of (1) for outcome variables not reported in Figure 12, alongside estimates of (21) that also include the expected counterfactual exposure to each temperature bin as controls. Daily temperature realizations in each county used to construct each temperature bins are from the ERA-5 Land dataset. In Panel (a), the unit of observation is a age group-county-year (1970-2002) and the outcome variable is the mortality rate for each age group, measured using the CDC mortality files. The age groups are constructed as ages < 1, 1-44, 45-64, and > 65 following Deschênes and Greenstone (2011). In Panels (b) and (c), the unit of observation is a county-year (1985-2006) and the outcome variable is log of soybean yield (b) and log of wheat yield (c), measured using NASS Quickstats. In Panel (d), the unit of observation is a county-month (1970-2009) and the outcome variable is nonviolent crime (i.e., robbery, larceny) measured using data from the FBI. 95% confidence intervals are reported and standard errors are clustered by county.



(a) Mortality Rate (All Ages, Stacked)

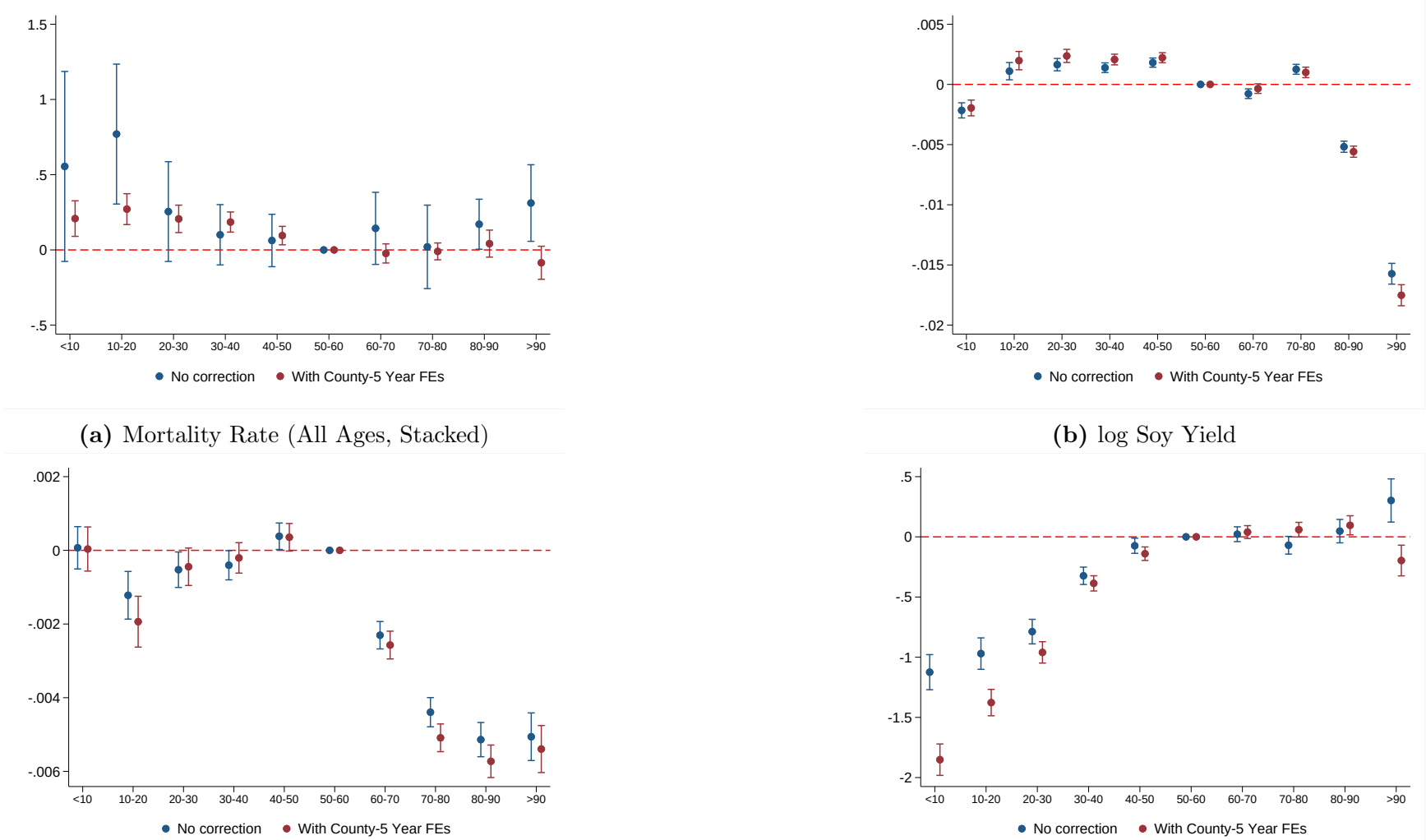
(b) log Soy Yield

(c) log Wheat Yield

(d) Nonviolent Crime

Figure A.14: Binning Bias with Real Outcome Data: Additional Outcomes (County Specific Linear Trends)

Notes: This figure reports estimates of (1) for outcome variables not reported in Figure A.11, alongside estimates of (21) that also include county specific linear trends as controls. Daily temperature realizations in each county used to construct each temperature bins are from the ERA-5 Land dataset. In Panel (a), the unit of observation is a age group-county-year (1970-2002) and the outcome variable is the mortality rate for each age group, measured using the CDC mortality files. The age groups are constructed as ages < 1, 1-44, 45-64, and > 65 following Deschênes and Greenstone (2011). In Panels (b) and (c), the unit of observation is a county-year (1985-2006) and the outcome variable is log of soybean yield (b) and log of wheat yield (c), measured using NASS Quickstats. In Panel (d), the unit of observation is a county-month (1970-2009) and the outcome variable is nonviolent crime (i.e., robbery, larceny) measured using data from the FBI. 95% confidence intervals are reported and standard errors are clustered by county.



(a) Mortality Rate (All Ages, Stacked)

(b) log Soy Yield

(c) log Wheat Yield

(d) Nonviolent Crime

Figure A.15: Binning Bias with Real Outcome Data: Additional Outcomes (County-5 Year Fixed Effects)

Notes: This figure reports estimates of (1) for outcome variables not reported in Figure A.12, alongside estimates of (21) that also include county-5 year fixed effects as controls. Daily temperature realizations in each county used to construct each temperature bins are from the ERA-5 Land dataset. In Panel (a), the unit of observation is a age group-county-year (1970-2002) and the outcome variable is the mortality rate for each age group, measured using the CDC mortality files. The age groups are constructed as ages < 1, 1-44, 45-64, and > 65 following Deschênes and Greenstone (2011). In Panels (b) and (c), the unit of observation is a county-year (1985-2006) and the outcome variable is log of soybean yield (b) and log of wheat yield (c), measured using NASS Quickstats. In Panel (d), the unit of observation is a county-month (1970-2009) and the outcome variable is nonviolent crime (i.e., robbery, larceny) measured using data from the FBI. 95% confidence intervals are reported and standard errors are clustered by county.

Table A.1: Comparison of Different `cftemp` Methods (ERA 5 (1970-2019))

Dataset	Method	Coef	t-Stat	$\omega_{C \text{ or } H}$	$\sigma^2_{e_{C \text{ or } H}}$	ω_Y	Bias
ERA 5 (1970-2019) $\sigma^2_{T_0} = 84.86$	*Under 10 Bin*						
	No correction	4.69935 [3.80523, 5.6083]	9.19458 [7.4216, 11.11126]	.0058857 [.0056941, .0060774]	.0017149	1.0008 [.96583, 1.03348]	3.35236 [3.23523, 3.46182]
	State-Year Fixed Effects	1.88201 [-.4247, 4.07439]	1.75086 [-.38996, 3.74433]	.0005417 [.0004121, .0006713]	.0007959	.12241 [.11016, .1347]	.03996 [.03597, .04397]
	With 3 Lags	3.36309 [2.27861, 4.38841]	6.1794 [4.19969, 8.08953]	.0041661 [.0039682, .0043641]	.0018307	.83614 [.80168, .87139]	2.0432 [1.95899, 2.12933]
	County-Specific Linear Trends	.01339 [-1.01502, 1.02644]	.02541 [-1.98966, 2.02286]	1.52e-18 [9.91e-19, 2.05e-18]	1.30e-32	0 [0, 0]	0 [0, 0]
	County-5 Year Fixed Effects	.38822 [-.71123, 1.42908]	.69757 [-1.29829, 2.58425]	.0003486 [.0003304, .0003668]	.0000155	.00974 [.00624, .01324]	.00211 [.00135, .00287]
	Counterfactual Temp. Control without Bayes	.2641 [-.75707, 1.30544]	.52361 [-1.48777, 2.5788]	.0002517 [.0001547, .0003487]	.0004394	.31611 [.29577, .33555]	.04929 [.04612, .05232]
	Counterfactual Temp. Control	.06665 [-.95401, 1.1317]	.13149 [-1.89087, 2.19296]	-6.60e-06 [-.0001174, .0001042]	.0005889	.24296 [.22567, .26012]	-.00097 [-.00104, -.0009]
	Chebyshev	.23994 [-.77675, 1.27405]	.46297 [-1.5282, 2.54045]	.0001891 [.0000901, .0002882]	.0004579	.29047 [.27004, .30838]	.03403 [.03164, .03613]
	Over 90 Bin						
	No correction	4.26788 [3.4888, 5.05203]	10.24993 [8.31996, 12.30084]	.0091595 [.0086446, .0096745]	.0123849	1.0008 [.96583, 1.03348]	5.81209 [5.60902, 6.00186]
	State-Year Fixed Effects	1.70329 [.34653, 3.00607]	2.52727 [.53288, 4.49999]	.0013078 [.0009482, .0016674]	.0061324	.12241 [.11016, .1347]	.11244 [.10119, .12373]
	With 3 Lags	2.963 [2.14093, 3.75624]	7.10831 [5.10124, 9.06581]	.0070156 [.0066191, .0074122]	.0073426	.83614 [.80168, .87139]	3.99881 [3.834, 4.16739]
	County-Specific Linear Trends	-.02138 [-.82123, .81351]	-.04966 [-1.91309, 1.90497]	-9.17e-18 [-9.92e-18, -8.43e-18]	2.61e-32	0 [0, 0]	0 [0, 0]
	County-5 Year Fixed Effects	-.02427 [-1.01834, .89936]	-.04951 [-2.09792, 1.88826]	3.67e-06 [-.0000142, .0000215]	.0000149	.00974 [.00624, .01324]	.00003 [.00002, .00004]
	Counterfactual Temp. Control without Bayes	.1789 [-.63344, 1.02585]	.42668 [-1.48377, 2.45267]	.0003218 [.0001806, .0004629]	.0009305	.31611 [.29577, .33555]	.07663 [.0717, .08134]
	Counterfactual Temp. Control	.03298 [-.78326, .86426]	.07682 [-1.86125, 2.05716]	.0000737 [-.0000674, .0002147]	.0009549	.24296 [.22567, .26012]	.01314 [.0122, .01406]
	Chebyshev	-.00944 [-.82895, .83396]	-.02265 [-1.92314, 1.99029]	.0000547 [-.0000693, .0001788]	.000719	.29047 [.27004, .30838]	.012 [.01116, .01274]

Notes: All temperatures are in °F. The sample period is 1970-2019 for ERA Land 5. All columns report statistics related to our main simulation exercise. Panel A reports statistic related to the under-10 degree bin and Panel B reports statistics related to the over-90 degree bin. All terms are as defined in Section 2 and the bias formula is presented in equations (12) and (13).

Table A.2: Adaptation Specification with Bias Formula

Dataset	County	Coef	t-Stat	$\omega_{C/H}$	$\sigma^2_{e_{C/H}}$	ω_Y	Bias
ERA 5 (1970-2019) $\sigma^2_{T_{0C}} = 23.2$ $\sigma^2_{T_{0H}} = 23.05$	*Under 10 Bin*						
	Cold Counties	1.15085 [.02995, 2.27833]	1.96219 [.0515, 3.92957]	.0063107 [.0056687, .0069526]	.0028002	1.00083 [.91083, 1.09079]	.71705 [.65257, .7815]
	Hot Counties	14.71374 [8.46642, 21.51152]	4.34773 [2.5078, 6.3566]	.002728 [.0025359, .0029201]	.0002188	1.00192 [.90989, 1.08813]	28.62393 [25.99485, 31.08701]
	Over 90 Bin						
	Cold Counties	2.56201 [-1.17933, 6.46276]	1.3624 [-.63144, 3.4819]	.0016414 [.0014252, .0018576]	.0003176	1.00083 [.91083, 1.09079]	6.42263 [5.8451, 6.99991]
	Hot Counties	1.52085 [.63091, 2.44204]	3.28932 [1.36722, 5.26493]	.0168816 [.0149002, .0188631]	.0232782	1.00192 [.90989, 1.08813]	1.60316 [1.45591, 1.74112]

Notes: All temperatures are in °F. The sample period is 1970-2019 for ERA Land 5. The first column reports the coefficient from running 1,000 simulations as described in Section 4.2 with three bins – hot, medium, and low – separately for cold and hot counties.