I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16253

Fatal Errors: The Mortality Value of Accurate Weather Forecasts

Jeffrey G. Shrader
Laura Bakkensen
Derek Lemoine

JUNE 2023

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16253

Fatal Errors: The Mortality Value of Accurate Weather Forecasts

Jeffrey G. Shrader
Columbia University and IZA

Laura Bakkensen
University of Arizona

Derek Lemoine
University of Arizona, NBER and CEPR

JUNE 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Fatal Errors: The Mortality Value of Accurate Weather Forecasts*

We provide the first revealed preference estimates of the benefits of routine weather forecasts. The benefits come from how people use advance information to reduce mortality from heat and cold. Theoretically, more accurate forecasts reduce mortality if and only if mortality risk is convex in forecast errors. We test for such convexity using data on the universe of mortality events and weather forecasts for a twelve-year period in the U.S. Results show that erroneously mild forecasts increase mortality whereas erroneously extreme forecasts do not reduce mortality. Making forecasts 50% more accurate would save 2,200 lives per year. The public would be willing to pay $112 billion to make forecasts 50% more accurate over the remainder of the century, of which $22 billion reflects how forecasts facilitate adaptation to climate change.

JEL Classification: D83, I12, Q51

Keywords: weather forecasts, information provision, mortality, climate change

Corresponding author:
Jeffrey G. Shrader
School of International and Public Affairs
Columbia University
116th and Broadway
New York, NY 10027
USA

E-mail: jgs2103@columbia.edu

* Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2112. (CBDRB-FY22-P2112-R9939) We thank Michael Best, Matthew Gibson, Michael Greenstone, Matthew Neidell, and Wolfram Schlenker as well as seminar participants at the NBER Environment and Energy Economics Spring Meeting, ASSA Annual Meeting, and several university seminars for helpful comments and suggestions. Special thanks to Anouch Missirian for providing code and guidance on dealing with wind data. Gabriel Gonzalez Sutil provided excellent research assistance. All errors are our own.

1 Introduction

Routine weather forecasts are the product of a sophisticated scientific and policy effort requiring global cooperation (Benjamin et al., 2018). Data from around the world are continuously gathered and processed to produce forecasts that are disseminated to the public multiple times per day, largely for free. Despite their ubiquity and the effort that goes into their creation, surprisingly little is known about their value to the public. Understanding this value is important for at least two reasons. First, weather forecasts are a far-reaching information intervention that is particularly well-suited to offering empirical insights into how individuals use information.[1] Yet economists have only rarely used forecasts to study informational questions (e.g., Roll, 1984). Second, forecasts are a costly but potentially important policy tool. Globally, public funding for forecasting agencies exceeds $15 billion per year (Rogers and Tsirkunov, 2013),[2] and decisions about future investments in forecasts require information about the value of making forecasts more accurate.[3] Economists have studied the value of weather warnings (e.g., Craft, 1998) but have not valued improvements in weather forecasts.

We combine new theory and comprehensive data to provide the first revealed preference evidence of behavioral responses to forecasts and the value of improved forecasts. In particular, we examine how much more accurate forecasts help people avoid mortality from temperature. We do so by formally demonstrating that whether more accurate forecasts reduce mortality depends on the relationship between mortality risk and forecast errors. If mortality risk is convex in forecast errors, then improvements in forecast accuracy will reduce expected mortality. The convexity of mortality risk depends on how agents' chosen actions depend on whether they receive one forecast or another. Namely, risk is convex if actions are "appropriate" in the sense of seeking to suit the weather rather than being "protective" in the sense of monotonically reducing mortality as more action is taken. Because either model of adaptation is plausible, it is an empirical question whether or not more accurate forecasts reduce mortality.

[1] Economists have long been interested in the role of information in decision-making, with particular growth in experimental evaluations of information interventions. Many experimental information interventions find effects on subjective beliefs but not actions, a difference that may reflect small sample sizes (Haaland et al., 2023). We take advantage of the ubiquity of forecasts to measure how people respond to information in a higher-powered environment.

[2] Each year, the U.S. National Weather Service has a budget of around $1 billion, collects around 76 billion weather observations, and issues around 1.5 million forecasts (NOAA, 2021).

[3] This point as been emphasized by both national and international observers (Chapman, 1992, WMO et al., 2015). Weather forecasts have steadily improved since 1980 (Bauer et al., 2015). Many expect forecast skill to continue increasing in the near future as weather forecasting groups improve observational networks, data assimilation, modeling, and computing power (Toth and Buizza, 2019, Zhang et al., 2019). Many have emphasized the need for good estimates of the economic value of forecast improvements (e.g., Freebairn and Zillman, 2002, Pielke and Carbone, 2002, National Research Council, 2010, Katz and Lazo, 2011).

To estimate the effect of weather forecasts on mortality, we combine the universe of deaths reported by the Centers for Disease Control and Prevention (CDC) with daily weather and forecasts issued by the National Weather Service (NWS). We study the continental U.S. from 2005 through 2017 and focus on day-ahead forecasts of temperature. A large literature across economics, epidemiology, and other fields has shown that extreme temperatures are a major source of mortality,[4] the NWS says that one goal when issuing forecasts is to minimize loss of life,[5] and households say these short-run temperature forecasts are among the most important forecast products to them (Stratus Consulting Inc., 2002). Our regression framework accounts for potential location-specific and time-varying confounders as well as for the potential direct effects of temperature and other aspects of weather on mortality.

Across the full sample, we find that mortality risk is indeed convex in forecast errors. Reducing the standard deviation of forecast errors by 50% would save 2,200 lives per year. We find that mortality risk is especially convex in forecast errors on days with extreme heat. This convexity indicates that either adaptation is particularly responsive to forecasts during periods of extreme heat or adaptation is especially consequential for mortality in extreme heat. Climate change will make such days more common over the coming century. As a result, climate change increases the mortality benefit of improved forecasts to 2,400 lives saved annually by 2100. Short-run weather forecasts thereby facilitate adaptation to climate change.

We also formally show how to estimate agents' ex ante willingness to pay for more accurate forecasts, accounting for their costs of acting on forecasts. Related literature arrives at net value for climate-induced mortality by using cross-sectional variation in climate and mortality risk to empirically identify marginal benefits of adaptation, then using first-order conditions to equate those marginal benefits of adaptation to marginal costs of adaptation, and finally summing these adaptation costs with mortality benefits estimated from panel variation in weather (Carleton et al., 2022). Instead, we formally derive the change in an agent's value function from more accurate forecasts and we measure that change from panel variation in forecasts and weather, without relying on cross-sectional variation for identification.[6] We find that the net value of making forecasts 50% more accurate is $2.1 billion per year, or roughly 10% of the monetized mortality benefit. Climate change increases that value to $2.9

[4]See, for instance, Anderson and Bell (2009), Deschênes and Moretti (2009), Gasparrini et al. (2015), Barreca et al. (2016), Carleton et al. (2022).

[5]This goal is part of the NWS mission statement: "Provide weather, water and climate data, forecasts, warnings, and impact-based decision support services for the protection of life and property and enhancement of the national economy."

[6]An established literature theoretically analyzes the value of changes in mortality risk (e.g., Berger et al., 1987, Shogren and Crocker, 1991), which corresponds to valuing changes in temperature in our setting. One can view Carleton et al. (2022) as implementing equation (17) in Berger et al. (1987). We differ in analyzing and estimating the value of changes in uncertainty about a mortality risk.

billion per year by 2100, so that the present value of improving forecast accuracy by 50% over the remainder of the century is $112 billion.

Forecasts can affect mortality only if people take actions based on them. We also report direct evidence that people do indeed act on forecasts: electricity use responds to forecasts, time use responds to forecasts, estimated responses to forecasts vary cross-sectionally in intuitive ways, and, in a new survey conducted for this paper, college students state that they modify behavior in response to forecasts. This evidence and the robustness of our results to model specification suggest that our estimated mortality benefits do indeed capture actual use of forecasts.

Our results are the first revealed preference estimates of the benefits of routine weather forecasts.[7] Recent theoretical work emphasizes that short-run forecasts such as those studied here can be especially valuable for planning purposes (Millner and Heyen, 2021). Previous valuations of routine weather forecasts calibrated models of particular decision problems (e.g., Lave, 1963, Wilks, 1997, Richardson, 2000), tallied up the value of sectors judged to be sensitive to weather (e.g., National Research Council, 1998), or surveyed potential users (e.g., Haas and Rinkle, 1979, Stewart, 1997, Stratus Consulting Inc., 2002, Lazo et al., 2009).[8] Many authors in the forecasting literature have recognized that it would be ideal to find a market in which people reveal their value for forecasts with real bets but lament that such markets do not exist for publicly provided forecasts (e.g., Freebairn and Zillman, 2002, Letson et al., 2007, Morss et al., 2008, Katz and Lazo, 2011). We here infer that agents use and value forecasts by exploring how forecasts affect observed mortality.

Previous work has found that both hot and cold temperatures are associated with excess mortality (see footnote 4). The dominant methodology uses fixed effects to account for the average weather in each place over time and also across places within a period. This methodology seeks to isolate the consequences of as-good-as-random weather shocks, but shocks relative to average weather may not be shocks relative to expectations: the "unusual" weather that provides the identifying variation in fixed effects models may or may not be surprising weather relative to the forecasts that agents see.[9] Accurately forecasted weather shocks can have very different implications from inaccurately forecasted weather shocks if people act on their information about coming weather. It is important to disen-

[7]Seasonal (i.e., multi-month) weather forecasts have been subject to additional study (e.g., Meza et al., 2008, Rosenzweig and Udry, 2019, Shrader, 2023). Shorter-run forecasts of weather only days ahead are much more accurate, widely produced, and widely used than seasonal forecasts. Some have shown that traders of weather derivatives attend to short-run forecasts like those studied here (e.g., Dorfleitner and Wimmer, 2010, Schlenker and Taylor, 2021), but they do not relate these forecasts to real outcomes.

[8]There have also been randomized controlled trials of forecast provision in developing country agricultural contexts (Yegbemey et al., 2023).

[9]Indeed, we show in Section 3 that forecasts substantially outperform average weather at predicting coming weather.

tangle these effects when assessing policy responses to extreme temperatures and also when extrapolating to the effects of climate change. First, the proper policy response depends on whether mortality is driven by a forecasting failure or by an extreme temperature. Second, estimated effects are typically driven by extreme temperatures. The most extreme realized temperatures could tend to exceed their forecasts, but these same temperatures may nonetheless be well-forecasted when they occur regularly with climate change. Reinforcing both concerns, we find that surprising extreme heat is deadly but forecasted extreme heat is not as deadly.

Our study shows that one of the most pervasive, prominent, and complex informational interventions undertaken by governments does generate substantial value for agents in the economy. Similar evidence is surprisingly rare. Governments have undertaken large-scale informational interventions for many decades, but many of these have been judged of dubious consequence (e.g., Adler and Pittle, 1983, Viscusi, 1989). Previous work documents that information provision can help people avoid air pollution (Neidell, 2009, Barwick et al., 2019, Wang and Zhang, 2022), enroll in retirement plans (Duflo and Saez, 2003), and choose schools (Hastings and Weinstein, 2008). We document the consequences for mortality of an intervention that is more thoroughgoing and consistent than any of these. Previous work shows that access to information moderates the effect of temperature on mortality during the Great Depression in the U.S. (Fishback et al., 2011), and concurrent work shows that the typical accuracy of forecasts affects the sensitivity of labor supply to forecasts of temperature in China (Song, 2022). Previous work also studies the effects of forecasting extreme weather events (e.g., Craft, 1998, Bakkensen, 2016, Miller, 2018, Weinberger et al., 2018, Kruttli et al., 2019, Molina and Rudik, 2022, Anand, 2022). We study the effects of the forecasts that underpin several important extreme weather warnings. Our headline value incorporates the effects of any extreme weather warnings that are based on temperature forecasts, but we do show that extreme heat warnings do not contribute much to our estimates.

Section 2 theoretically analyzes how more accurate forecasts can reduce expected mortality. Section 3 describes the data used for the empirical analysis. Section 4 reports results on the relationship between mortality and forecast errors. Section 5 analyzes mechanisms. Section 6 calculates the mortality benefit of improved forecasts, willingness to pay for improved forecasts, and how each depends on climate change. Section 7 concludes.

2 Theoretical Analysis

We extend the workhorse single-period model of the value of a statistical life (VSL), the marginal rate of substitution between money and the risk of sudden mortality, to allow for mortality due to temperature and for adaptation based on forecasts. This model originates

with Drèze (1962) and Jones-Lee (1974) and has been extensively applied in the literature (see Viscusi, 1993, Andersson and Treich, 2011).

An individual's indirect utility over wealth is $u(\cdot)$ when alive and $v(\cdot)$ when dead. Per convention, assume that $u(\cdot) > v(\cdot)$, $u'(\cdot) \geq v'(\cdot) \geq 0$, and $u''(\cdot) < 0$. The individual's hazard of death following temperature T is $h(T, A) \geq 0$, where $h_{TT} \geq 0$ and $h_{AT} \neq 0$ (with subscripts indicating partial derivatives). A represents actions chosen with knowledge of a forecast f of temperature. These actions constitute "ex ante" or "anticipatory" adaptation.[10] Forecasts are unbiased: $E[T|f] = f$.[11] Actions cost $C(A)$, where, for convenience, $C(\cdot)$ is differentiable and $C''(\cdot) \geq 0$. Initial wealth is w, so that end-of-period wealth is $w - C(A)$.

The individual chooses actions to maximize expected indirect utility

$$V(f) = \max_A E_{T|f}\Big[[1 - h(T, A)]\, u(w - C(A)) + h(T, A)\, v(w - C(A))\Big], \tag{1}$$

where $E_{T|f}$ indicates expectations over temperature given forecast f. The first-order condition is

$$\begin{aligned} 0 = &- C'(A)\, u'(w - C(A)) \\ &+ E_{T|f}[h(T, A)]\, C'(A) \Big(u'(w - C(A)) - v'(w - C(A))\Big) \\ &- E_{T|f}[h_A(T, A)]\Big(u(w - C(A)) - v(w - C(A))\Big). \end{aligned} \tag{2}$$

This equation implicitly defines optimal actions $A^*(f)$, with the second-order condition discussed in Appendix B.1. Now consider adding a small daily mortality risk τ to $h(T, A)$. The VSL is willingness to accept the small risk. Substituting $A^*(f)$ into (1), totally differentiating while holding V constant, and using (2), we find:

$$\begin{aligned} VSL(f) &\triangleq \left.\frac{\mathrm{d}w}{\mathrm{d}\tau}\right|_{\tau=0} \\ &= \frac{u(w - C(A^*(f))) - v(w - C(A^*(f)))}{[1 - E_{T|f}[h(T, A^*(f))]]\, u'(w - C(A^*(f))) + E_{T|f}[h(T, A^*(f))]\, v'(w - C(A^*(f)))}. \end{aligned} \tag{3}$$

[10] This representation subsumes any "ex post" or "reactive" adaptation (i.e., adaptation chosen based on knowledge of realized temperature) into the temperature argument of the hazard function.

[11] If we permitted nonzero expected bias, then the below expressions would survive with derivatives evaluated at the new conditional expectation. The bias is small in our empirical application.

Around an optimum, the first-order condition (2) becomes

$$C'(A^*(f)) = -E_{T|f}[h_A(T, A^*(f))]\, VSL(f). \tag{4}$$

Optimal adaptation equates its marginal cost to its marginal benefit, measured as the expected reduction in mortality risk times the value of reduced mortality risk.

The following assumption will suit our empirical application, in which agents value their lives highly and have access to high-quality weather forecasts:

Assumption 1 (Small Errors and Valuable Lives) *For some forecast f and for ϵ positive but small, $C'(A^*(f))/VSL(f) \leq \epsilon$ and $Var[T|f] \leq \epsilon$.*

Under this assumption, adaptation has a small marginal effect on mortality risk around agents' optimal choices:

Lemma 1 (Small Marginal Adaptation Effects) *Under Assumption 1,*

$$\lim_{\epsilon \to 0} h_A(f, A^*(f)) = 0.$$

Proof. Second-order approximating $h_A(\cdot, A^*(f))$ around $T = f$,

$$E_{T|f}[h_A(T, A^*(f))] \approx h_A(f, A^*(f)) + \frac{1}{2} h_{ATT}(f, A^*(f)) Var[T|f], \tag{5}$$

with higher-order terms vanishing as $Var[T|f]$ becomes small. $Var[T|f] < \epsilon$ implies

$$\lim_{\epsilon \to 0} E_{T|f}[h_A(T, A^*(f))] = \lim_{\epsilon \to 0} h_A(f, A^*(f)).$$

From (4), $C'(A^*(f))/VSL(f) < \epsilon$ implies that $\lim_{\epsilon \to 0} E_{T|f}[h_A(T, A^*(f))] = 0$. ■

Agents adapt to nearly the point where their adaptations cease affecting mortality risk in the event that their forecast turns out to be accurate.

Now consider the average mortality risk induced by forecast errors, which determines lives saved from accurate forecasts. We observe days with temperature T and ask how average mortality risk on these days would have changed if forecasts had been more accurate. Let the forecast error on a day with measured temperature T be $e \triangleq f - T$, so that the optimal adaptation action is $A^*(T + e)$. Average mortality risk on a day with temperature T is $\bar{h}(T) \triangleq E_{e|T}[h(T, A^*(T + e))]$, where we average over the possible forecast errors for a day

with temperature T. Second-order approximating around $e = 0$,

$$\bar{h}(T) \approx h(T, A^*(T)) + \frac{1}{2} \underbrace{\left(h_A(T, A^*(T))\, A^{*\prime\prime}(T) + h_{AA}(T, A^*(T))\, [A^{*\prime}(T)]^2\right)}_{\mathrm{d}^2 h(T, A^*(T+e))/\,\mathrm{d}e^2|_{e=0}} Var[e|T]. \quad (6)$$

Let the number of days with temperature T during a year be $n(T)$. The additional mortality risk over the year from making forecasts less accurate is

$$\frac{\mathrm{d}\bar{h}(T)}{\mathrm{d}Var[e|T]} n(T) \approx \underbrace{\frac{1}{2} \left.\frac{\mathrm{d}^2 h(T, A^*(T+e))}{\mathrm{d}e^2}\right|_{e=0}}_{\text{ex post increase in errors}} n(T) \quad (7)$$
$$+ \underbrace{\left(h_A(T, A^*(T)) + \frac{1}{2} \left.\frac{\mathrm{d}^3 h(T, A^*(T+e))}{\mathrm{d}e^2\, \mathrm{d}A}\right|_{e=0} Var[e|T]\right) \frac{\mathrm{d}A^*(T)}{\mathrm{d}Var[e|T]}}_{\text{effect of uncertainty on ex ante adaptation}} n(T).$$

The first line on the right-hand side gives the mortality incurred once we average over the ex post mistakes. It depends on the convexity of mortality risk in forecast errors, which in turn depends on how agents react to different forecasts. The second line gives the change in mortality due to agents' rational expectations of forecast accuracy. From (2), A^* will depend on the variance of forecast errors unless mortality risk is linear in temperature. Agents who understand that a forecast's accuracy has changed may choose different actions that in turn generate different mortality risks.

We will be able to credibly identify the ex post error term on the first line of (7) from random variation in forecast errors. The following proposition establishes that the ex post error term is likely to be the important term in our empirical application:[12]

Proposition 1 (Effect of Forecast Accuracy on Average Mortality) *If Assumption 1 holds at $f = T$, then*

$$\lim_{\epsilon \to 0} \frac{\mathrm{d}\bar{h}(T)}{\mathrm{d}Var[e|T]} n(T) = \lim_{\epsilon \to 0} \frac{1}{2} \left.\frac{\mathrm{d}^2 h(T, A^*(T+e))}{\mathrm{d}e^2}\right|_{e=0} n(T).$$

Proof. Follows from (7) and Lemma 1. ■

If we estimate that mortality risk is convex in forecast errors, then we can conclude that forecasts would reduce average mortality risk.

We broadly classify agents' adaptation environments based on how adaptation affects

[12] Proposition 1 also holds if, instead of Assumption 1, we assumed that $|h_{ATT}| \leq \epsilon$, because then $\lim_{\epsilon \to 0} \mathrm{d}A^*(T)/\,\mathrm{d}Var[e|T] = 0$ in (7). For instance, Proposition 1 would hold if we assumed quadratic $h(\cdot, \cdot)$ instead of Assumption 1.

mortality risk when forecasts turn out to be inaccurate:

Definition 1 (Adaptation Environment) Adaptation is *appropriate* if, for forecasts f in a neighborhood of T, $\mathrm{d}h(T, A^*(f))/\,\mathrm{d}f < 0$ for $f < T$ and $\mathrm{d}h(T, A^*(f))/\,\mathrm{d}f > 0$ for $f > T$. Adaptation is *protective* otherwise.

When adaptation is appropriate, an action's mortality risk is minimized when the action is based on a forecast that turns out to be accurate:

Corollary 1 (Forecast Errors that Minimize Mortality) *If Assumption 1 holds at $f = T$ with ϵ arbitrarily small and adaptation is appropriate, then realized mortality on a day with temperature T is locally minimized when there is no forecast error.*

Proof. From Definition 1, $\mathrm{d}^2h(T, A^*(T+e))/\,\mathrm{d}e^2\big|_{e=0}$ is strictly positive when adaptation is appropriate, and $\mathrm{d}^2h(T, A^*(T+e))/\,\mathrm{d}e^2\big|_{e=0} > 0$ is a sufficient condition for mortality risk to be minimized at $e = 0$ in the case that the first-order necessary condition $\mathrm{d}h(T, A^*(T+e))/\,\mathrm{d}e\big|_{e=0} = 0$ holds at $e = 0$. Observe that

$$\left.\frac{\mathrm{d}h(T, A^*(T+e))}{\mathrm{d}e}\right|_{e=0} = h_A(T, A^*(T))\, A^{*\prime}(T).$$

Applying Lemma 1, $\lim_{\epsilon\to 0} \mathrm{d}h(T, A^*(T+e))/\,\mathrm{d}e\big|_{e=0} = 0$. Therefore the first-order condition for errors that locally minimize mortality risk holds at $e = 0$ as ϵ goes to 0. ■

Because both positive and negative forecast errors locally increase mortality risk when adaptation risk is appropriate, reducing the variance of forecast errors reduces average mortality risk when adaptation is appropriate:

Corollary 2 (Form of Adaptation Determines Effect of Forecast Accuracy) *If Assumption 1 holds at $f = T$ with ϵ arbitrarily small, then reducing the variance of forecast errors on a day with temperature T reduces average mortality if and only if adaptation is appropriate.*

Proof. From (6) and Lemma 1, $\mathrm{d}^2h(T, A^*(T+e))/\,\mathrm{d}e^2\big|_{e=0} > 0$ if and only if $h_{AA}(T, A^*(T)) > 0$. Because $\mathrm{d}h/\,\mathrm{d}e\big|_{e=0} = 0$ from Lemma 1, $h_{AA}(T, A^*(T)) > 0$ implies that $\mathrm{d}h/\,\mathrm{d}e$ changes sign from negative to positive around $e = 0$. The result follows from Proposition 1 and Definition 1. ■

When Assumption 1 holds, the condition from Proposition 1 for mortality risk to increase in the variance of forecast errors is equivalent to adaptation being appropriate.

When adaptation is appropriate, adaptation is targeted to particular temperatures and is less effective if temperatures are higher or lower than the target.[13] Formally, appropriate adaptation could have $h(T, A) \propto (T - A)^2$, so that mortality increases away from $A = T$. In such an environment, symmetric forecast mistakes can increase mortality on average. We label other cases as protective adaptation. For instance, mistakenly undertaking too much adaptation may reduce mortality risk compared to a case in which the realized temperature had been forecasted accurately. Formally, protective adaptation could have the hazard decrease monotonically in A, as when $h(T, A) \propto T(T - A)$ for $A \leq T$. In such an environment, symmetric forecast mistakes can fail to affect (or can even reduce) mortality on average. It is not *a priori* obvious which type of adaptation predominates in the real world. It is ultimately an empirical question whether more accurate forecasts in fact reduce average mortality. And from Proposition 1, we can estimate whether they do so by estimating the relationship between mortality risk and forecast errors.

We have thus far followed much other literature (e.g., Deschênes and Moretti, 2009, Barreca et al., 2016) in studying the mortality consequences of weather without netting out adaptation costs. Now consider each agent's willingness to pay for more accurate forecasts, net of adaptation costs. Consider an agent who knows the temperature will be T on $n(T)$ days in the year and values the accuracy of the forecasts she receives in advance of those days. Let the conditional probability of forecast f be $p(f|T)$. Expected value on a day with temperature T is

$$\bar{V}(T) \triangleq \int V(f)\, p(f|T)\, \mathrm{d}f.$$

Denote the agent's willingness to pay for a marginally more accurate forecast on days with temperature T as $WTP(T)$:

$$WTP(T) \triangleq -\frac{\mathrm{d}\bar{V}(T)/\,\mathrm{d}Var[e|T]}{\mathrm{d}\bar{V}(T)/\,\mathrm{d}w} n(T). \tag{8}$$

Assume that observing forecasts similar to the accurate forecast does not change the agent's perception of accuracy:

Assumption 2 (Locally Constant Forecast Accuracy) *$Var[e|f]$ is constant around forecast $f = T$.*

We now value an improvement in forecast accuracy on days with temperature T:

Proposition 2 (Agent's Willingness to Pay for More Accurate Forecasts) *If Assump-*

[13] Appendix B.2 shows that we learn that $\lim_{\epsilon \to 0} h_{AA}(T, A^*(T)) > 0$ if we estimate that adaptation is appropriate and Assumption 1 holds.

tions 1 and 2 hold, then

$$\lim_{\epsilon \to 0} WTP(T) = \lim_{\epsilon \to 0} \frac{1}{2} VSL(T) \underbrace{\left[\left. \frac{\partial^2 h(T, A^*(T+e))}{\partial T^2} \right|_{e=0} + \frac{\mathrm{d}}{\mathrm{d}e} \left. \frac{\partial h(T, A^*(T+e))}{\partial T} \right|_{e=0} \right]}_{\mathrm{d}h_T(T, A^*(T+e))/\,\mathrm{d}T} n(T). \tag{9}$$

Proof. See Appendix B.3. ■

The possibility of forecast errors makes the agent uncertain about what forecasts she will receive over the year. This uncertainty is costly if and only if $V(f)$ is concave around $f = T$ (equation (A-2)). The proof shows that $V''(f)$ depends on the total derivative of h_T with respect to T, which comprises the two terms in square brackets in (9).

The first term captures the change in the agent's expected mortality risk due to their reduced uncertainty about realized temperature. Via Jensen's inequality, expected mortality is greater when mortality risk is convex in temperature (i.e., when $h_{TT} > 0$). Therefore this first term generates positive willingness to pay when $h_{TT} > 0$.

The second term typically reduces agents' willingness to pay for more accurate forecasts because agents' responses to forecasts tend to make mortality risk less convex in temperature forecasts than in realized temperature (see Appendix B.3). It scales with h_{AT} and with $A^{*\prime}$, so that it becomes important when the effectiveness of actions depends on temperature and actions are sensitive to forecasts. Under general conditions, agents choose larger A upon observing a higher forecast if and only if higher temperatures increase the marginal mortality benefit of A (i.e., iff $h_{AT} < 0$; see Appendix B.2). So actions respond to forecasts in a direction that reduces h_T under high forecasts and increases h_T under low forecasts. These responses push mortality risk to be less convex in forecasted temperature. For instance, if hotter weather makes adaptation more effective, then agents respond to forecasted hotter weather by doing more adaptation and thereby mitigating any increase in mortality risk. Anticipating these responses reduces agents' willingness to pay for more accurate forecasts.

One might be surprised that willingness to pay is not driven by the mortality consequences of inaccurate forecasts (i.e., that the $\mathrm{d}^2h/\,\mathrm{d}e^2$ term from Proposition 1 does not appear). This absence is a consequence of the envelope theorem. The envelope theorem reflects the assumption that agents optimize their actions (i.e., that the first-order condition holds at all forecasts). If not for the envelope theorem, $V'(f)$ would include $\mathrm{d}h/\,\mathrm{d}e$, in which case $V''(f)$ would include the $\mathrm{d}^2h/\,\mathrm{d}e^2$ term that Proposition 1 showed controls the average mortality benefit of more accurate forecasts. However, $\mathrm{d}h/\,\mathrm{d}e$ matters only via $A^{*\prime}$, which drops out of $V'(f)$ because optimizing agents' marginal costs of adapting offset any marginal benefit. As a result, the mortality benefit of advance information drops out of value calculations. But

the envelope theorem does not apply when differentiating $V'(f)$. Because $V'(f)$ includes h_T, $V''(f)$ includes the derivative of h_T with respect to forecasts and thus includes how forecast-driven actions affect the risk h_T due to realized weather (i.e., the second term in (9)). The envelope theorem eliminated the effects of actions from previous work that valued marginal changes in mortality risk (e.g., Berger et al., 1987, Shogren and Crocker, 1991, Carleton et al., 2022), but actions here do affect net value because we value changes in uncertainty about that mortality risk.[14]

3 Data

To estimate the effect of forecasts on mortality risk, we combine data on mortality events, realized temperature, and temperature forecasts. Details on datasets used for robustness and auxiliary analyses can be found in Appendix A.

3.1 Weather and Weather Forecasts

Our two primary explanatory variables are daily temperature and forecasted temperature. We focus on 1-day-ahead, daily minimum and maximum temperature point forecasts, from which we calculate daily average temperature by averaging the two measures. The NWS runs the forecasting model multiple model times during the day, with the most important runs at noon and midnight UTC. We use the noon UTC run because it is typically the one reported in morning news broadcasts. These are also the forecasts that are available at any time of day on the public NWS website, `weather.gov`.[15]

The forecasts are stored in the National Digital Forecast Database (NDFD), which was created in the early 2000s to standardize the processing, storage, and dissemination of weather forecasts in the U.S. (Glahn and Ruth, 2003). Meteorologists in different locations across the country, known as Weather Forecasting Offices (WFOs), work in shifts to produce forecasts for their local area, known as County Warning Areas (CWAs, mapped in

[14] In concurrent work, Molina and Rudik (2022) analyze how reducing the uncertainty in a given hurricane wind speed forecast affects the expected cost of that hurricane. In their analysis, the reduction in forecast uncertainty reduces the spread of realized wind speeds, so that they value changes in mortality risk. In contrast, we analyze willingness to pay for systematically more accurate forecasts, before the particular forecasts have been produced. Forecast accuracy here does not affect realized weather but instead affects the spread of the forecasts that agents receive in advance of that weather. As a result, the concavity of the value function in the observed forecast is critical in our analysis.

[15] People may see forecasts that differ from the NWS forecasts. For instance, apps, websites, or local media may slightly alter the official forecasts, in which case the NWS forecasts serve as a proxy for observed forecasts. Such alterations would cause concern only if they introduce bias that is correlated with the NWS forecasts. This type of bias has been seen historically, for instance before baseball games in 1890s New York (Raymond and Taylor, 2021) and for more recent Weather Channel rainfall forecasts that exaggerate the probability of rainfall (Bickel and Kim, 2008). Modern temperature forecasts—the object of study in this paper—do not typically exhibit such biases (Brooks et al., 1997).

Figure A10). The NDFD stores the forecasts on a consistent spatial grid with resolution of 2.5km or 5km, depending on the time period.

We use forecast data in the NDFD that contains both minimum and maximum temperature forecasts. These data begin on April 13, 2005. Roughly 5% of the county-day values are missing in the raw NDFD data (owing, for example, to data corruption in the NWS archives or to a missed forecast deadline).[16] We aggregate the forecasts to the county level by taking the population-weighted average, based on the 2010 population grids from CIESIN (2017).

For weather realizations, we use PRISM (Parameter-elevation Regressions on Independent Slopes) Climate Group data (PRISM Climate Group, 2004). PRISM combines weather station observations with an interpolation procedure that accounts for features such as elevation, weather inversions, rain shadows, and coastal proximity. The outputs are daily measures of weather on a consistent 4km resolution grid across the country. The PRISM data provide more consistent geographic coverage than raw weather station data.[17] We calculate a day's average realized temperature by averaging the day's minimum and maximum realized temperature. We aggregate the gridded measures to the county level using the same procedure as for forecasts.

We map the spatial variation in weather and compare weather and forecast values in Appendix Figure A11. The final weather dataset contains measures of daily minimum and maximum temperature as well as control variables for total daily rainfall and average dew point temperature for each continental U.S. county from April 13, 2005 to July 6, 2017.

3.2 Mortality

Our primary outcome is mortality. Mortality data come from the CDC's National Center for Health Statistics Multiple Cause of Death (MCOD) file. It contains records of all vital events that occurred in the U.S. from 2005 to 2017. We use the restricted access version of the dataset, which includes the day and county of each mortality event. From the set of all mortality events, we calculate county mortality rates per 100,000 people by dividing the total mortality each day by the county population in that year. Population figures are from the NIH Surveillance, Epidemiology, and End Results (SEER) Program. The race, sex, and cause of death for the decedent are also recorded and are discussed further in the heterogeneity analysis results. Decedent demographics are typically filled out by the funeral

[16] The results we report below are robust to estimating with a sample that interpolates the missing values based on the most recent available forecast.

[17] We use the original PRISM data rather than the version produced by Schlenker and Roberts (2006, 2009) so that we can also measure dewpoint temperature for humidity robustness checks. One of the advantages of the Schlenker-Roberts version of the data is greater temporal consistency. Over the relatively recent time series analyzed in this project, the difference between the original PRISM data and the Schlenker-Roberts data is minimal.

home director or a healthcare work, ideally based on input from next of kin (Ver Ploeg and Perrin, 2004).

3.3 Data Structure and Summary Statistics

The primary estimation sample consists of all non-missing observations of all-cause mortality, average temperature, total daily precipitation, average 1-day-ahead temperature forecasts,[18] and population shares for each county in the continental U.S. and each day from April 13, 2005 to July 6, 2017. Summary statistics for the sample are shown in Appendix Table A3. In the estimation sample, the average number of deaths reported for all causes per day across the U.S. is 2.3 per 100,000 people.

Forecasts tend to be close to correct on average. There is a slight cool bias to the forecasts overall of about -0.04°C. This bias is small relative to the standard deviation of forecast errors, which is just over 1°C.[19] Figure 1a shows the distribution of forecast errors within four different expected temperature ranges: cold (< 0°C), cool (0–15°C), warm (15–30°C), and hot (> 30°C). Across the expected temperature distribution, forecasts are essentially unbiased and are largely symmetric around zero with a bell-curve shape. The distributions are widest for the coldest temperatures and narrowest for the warmest temperatures.

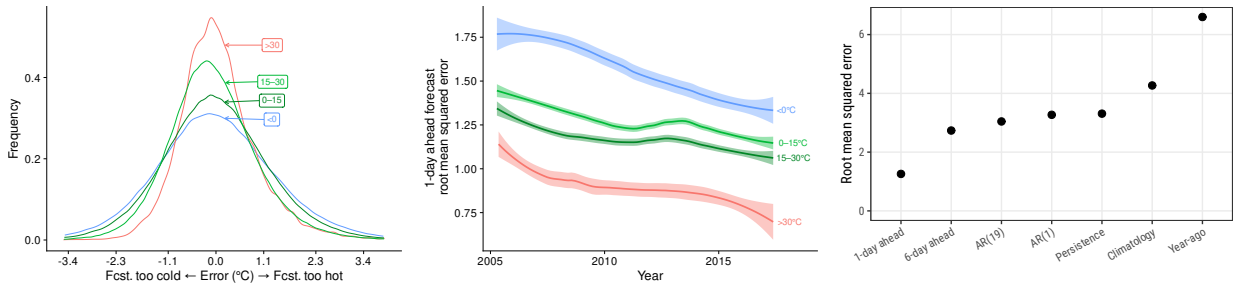
Figure 1b shows how 1-day-ahead forecast accuracy has evolved over our sample period. The lines in the figure are local linear regressions fit to the daily forecast root mean squared error (RMSE). Over the 12 years of the sample, forecast RMSE fell from about 1.37°C to about 1.08°C, an improvement of 27%. The forecasts have improved within all temperature ranges. The largest improvement occurred for hot temperatures, with a 56% improvement on an already high level of accuracy.

Figure 1c compares the forecast error for our sample to error from other forecasting methods. It shows the RMSE for the forecasts we use in our baseline analysis (the 1-day-ahead forecasts issued by the NWS) as well as six additional forecasts: the 6-day-ahead NWS forecast, an AR(19) forecast selected through information criteria, an AR(1) forecast using yesterday's value (estimated within sample), a persistence forecast based on yesterday's observation, a climatological forecast based on the average weather for that location and day, and the observation from one year ago. The 1-day-ahead forecast substantially outperforms

[18] In Section H we also analyze longer-horizon forecasts. Both longer- and shorter-horizon forecasts affect mortality, with decreasing marginal returns to horizon. Given the high correlation between forecasts of different horizons, the main results can be interpreted as the effect of forecasts of horizons between 1 and 6 days.

[19] See Myrick and Horel (2006) for an early analysis of forecast error and discussion of the creation of a routine verification system that can now be found at `https://sats.nws.noaa.gov/~verification/ndfd/`. Myrick and Horel (2006) emphasize how terrain features like rapid changes in elevation can affect forecast error. In our estimation sample, we include location fixed effects to mitigate bias from these types of differences.

Figure 1: Forecast Errors and Comparison With Alternatives



(a) Forecast Error Density (b) Accuracy (c) NWS fcst. vs. alternatives

Notes: Panel (a) shows empirical distributions of errors from the 1-day-ahead NWS forecast within four expected temperature ranges. The values are weighted by annual, county-level population, and distributions are truncated at the 0.025 and 99.75 percentiles. The x-axis tick marks are at standard deviations of the overall distribution of forecast errors. Panel (b) shows the trend in 1-day-ahead forecast root mean squared error (RMSE) over the sample period. Each line is a local linear regression fit to the daily, national average RMSE within 4 expected temperature ranges. Panel (c) shows the RMSE for the 1-day-ahead and 6-day-ahead NWS forecasts compared to alternatives: an AR(19) forecast selected through information criteria, an AR(1) forecast using yesterday's observation, a persistence forecast based on yesterday's observation, the average value for that location and day (the "forecast" when using location-time fixed effects), and last year's value for that location.

even the best-performing non-NWS competitor, suggesting that agents should use the day-ahead NWS forecasts that we study over other plausible alternatives.[20]

4 Identifying Behavior by Estimating the Relationship Between Mortality and Forecast Errors

Section 2 showed that the shape of the relationship between mortality risk and forecast errors is informative about the types of adaptation that people undertake and that the convexity of the relationship determines the mortality benefit of improving forecast accuracy. So the empirical analysis begins by estimating the shape of this relationship.

[20] Conventional weather-mortality regressions do implicitly account for the climatological forecast via location fixed effects, but Figure 1c shows that its RMSE is four times larger than the official forecast's RMSE. Conventional weather-mortality regressions therefore do not capture all of the information available to an agent one (or even six) days ahead of a weather realization.

4.1 Estimating Equation and Empirical Strategy

To assess behavior, we estimate the following equation, which allows for flexible but powerful tests of non-monotonicity in the effect of forecast errors on mortality:

$$y_{ct} = \sum_{\ell=0}^{L} \left[\sum_{j=1}^{J} \mathbf{1}\{f_{c,t-\ell} \in B_j\} \left(\beta_{10,\ell,j} + \beta_{11,\ell,j} e^{low}_{j,c,t-\ell} + \beta_{12,\ell,j} e^{high}_{j,c,t-\ell} + \beta_{13,\ell,j} \mathbf{1}\{e_{c,t-\ell} > \tilde{e}_j\} \right) \right] + \sum_{\ell=0}^{L} \left[g_{1,\ell}(T_{c,t-\ell}; \xi_{11}) + g_{2,\ell}(\text{prec}_{c,t-\ell}; \xi_{12}) \right] + X_{ct}\gamma_1 + \alpha_{1,cm} + \rho_{1,t} + \varepsilon_{1,ct}, \tag{10}$$

where $\mathbf{1}$ is an indicator function. The dependent variable y_{ct} is the daily mortality rate in county c on day t. The primary right-hand side variables are realized temperature T_{ct} and forecast error $e_{ct} \triangleq f_{ct} - T_{ct}$ (based on the one-day-ahead forecast f_{ct} of temperature).[21] In one specification that allows for more flexibility, we bin forecasted temperature every 5°C from 0°C to 30°C, for $J = 8$ bins in total, and in another specification that allows for clearer presentation, we bin every 15°C from 0°C to 30°C, for $J = 4$ bins in total.

The theoretical model shows that our estimating equation needs to be flexible enough to capture a potentially non-monotonic relationship between mortality and forecast errors. In particular, we want to classify behavior based on potential non-monotonicities without yet invoking Assumption 1. Non-monotonic or "U-shaped" relationships are sometimes tested using quadratic specifications, but this approach can lead to spuriously finding a U shape (Kostyshak, 2017). Simonsohn (2018) proposes a test to detect non-monotonicities that avoids false positives that can occur with quadratic specifications yet maintains power to detect true effects. Simonsohn's two-line approach uses an interrupted regression to estimate distinct slopes on either side of a breakpoint, $\tilde{e}$.[22] Those slopes are the coefficients $\beta_{11,\ell,j}$ and $\beta_{12,\ell,j}$ on

$$e^{low}_{j,c,t-\ell} \triangleq (\tilde{e}_j - e_{c,t-\ell})\mathbf{1}\{e_{c,t-\ell} \leq \tilde{e}_j\} \qquad \text{and} \qquad e^{high}_{j,c,t-\ell} \triangleq (e_{c,t-\ell} - \tilde{e}_j)\mathbf{1}\{e_{c,t-\ell} > \tilde{e}_j\}.$$

The coefficients can be interpreted as the marginal effect of a more erroneous forecast. If both coefficients are positive, then we conclude that adaptation is "appropriate" per Definition 1.

We estimate distributed lag models in order to account for dynamic effects such as temporal displacement of mortality. In the main results, we use 6 total lags of temperature,

[21] To reduce the influence of potential outliers, we Winsorize forecast errors at the 1% level.

[22] We set the breakpoint to the median forecast error within each expected temperature range. Section D assesses sensitivity to choice of breakpoints.

precipitation, and forecast errors ($L = 5$).[23] We focus on cumulative effects, defined as

$$\bar{\beta}_{i,j} \triangleq \sum_{\ell=0}^{5} \beta_{i,\ell,j}. \tag{11}$$

The other components of the estimating equation are controls. For temperature realizations, we estimate $g_1(T; \xi_1)$ semi-parametrically by discretizing temperature into 5°C bins. Flexibly controlling for temperature realizations ensures that our forecast error estimates can be interpreted as the effect of varying forecasts around a given realization. The function $g_2(\text{prec}; \xi_2)$ is specified as indicators for daily precipitation above the county's median. This control accounts for potentially correlated effects of rainfall. Both temperature and rainfall lags are included in the distributed lag model.

Date fixed effects (denoted $\rho_{1,t}$) remove any national, time-based confounders, including day-of-week effects, holidays, overall patterns in economic activity, changes in national policy, and large-scale weather patterns. County-by-month fixed effects ($\alpha_{1,cm}$) remove local seasonal patterns. These fixed effects also adjust for any time-invariant differences across locations including long-run climate, average medical care availability, economic conditions, and information provision and acquisition. X_{ct} includes county-month fixed effects interacted with linear time trends to account for any overall trends in county- or season-specific weather or mortality. Finally, we weight the regressions by annual population in the county to better estimate nationally representative values and to increase precision.

Standard errors are clustered at the CWA level. As discussed in Section 3, CWAs are collections of counties that receive weather forecasts from the same NWS forecast office. The counties are grouped based on technological considerations related to weather observation equipment and so that meteorologists can specialize in forecasting weather for particular areas of the country. The CWA is a natural level for clustering because its counties have similar weather and because its counties all receive forecasts from the same group of meteorologists. Econometrically, CWAs also have some further appealing features, particularly when compared to state-level clustering. First, there are 116 CWAs in the continental U.S., well above the typical rule-of-thumb for the minimum number of clusters (Cameron and Miller, 2015).[24] Second, state borders differ from CWAs in not being defined based on meteorological considerations, making it more plausible that meteorological data are approximately i.i.d. across CWAs than across states.

[23] Using 6 lags balances flexibly capturing dynamic effects with computational complexity. Previous epidemiological research shows that most adverse effects from heat appear within 3 days (Guo et al., 2014). We include up to 14 lags in robustness checks (Appendix C).

[24] Some counties fall into more than one CWA, and for those locations, we assign the county an identifier that is the combination of CWAs to it belongs. This results in 130 CWA ids in our sample. See Figure A2 for CWA boundaries.

For identification, we assume that forecast errors are as-good-as-randomly assigned within counties over time (i.e., sequential exogeneity), conditional on our other covariates. Formally, we assume that the expectation of ε_{ct} is zero, conditional on all contemporaneous and past values of our covariates. This assumption is more plausible when studying forecast errors than for a generic variable. If forecasting systems strive for accuracy, then historical atmospheric conditions will not confound forecast errors because the meteorologist takes them into account when formulating the forecast. In other words, forecast errors are surprises relative to the information available to the forecaster, so they cannot be confounded by anything inside that information set. The primary remaining potential threat to identification is unobserved aspects of realized *contemporaneous* weather that directly affect mortality and also affect forecast quality. To address this concern, we conduct robustness checks (see Section 4.3) that include additional controls for contemporaneous weather and atmospheric conditions, including humidity, wind, pollution, and other forecast errors. We find that effects are largely unchanged when these additional covariates are added.

4.2 Results

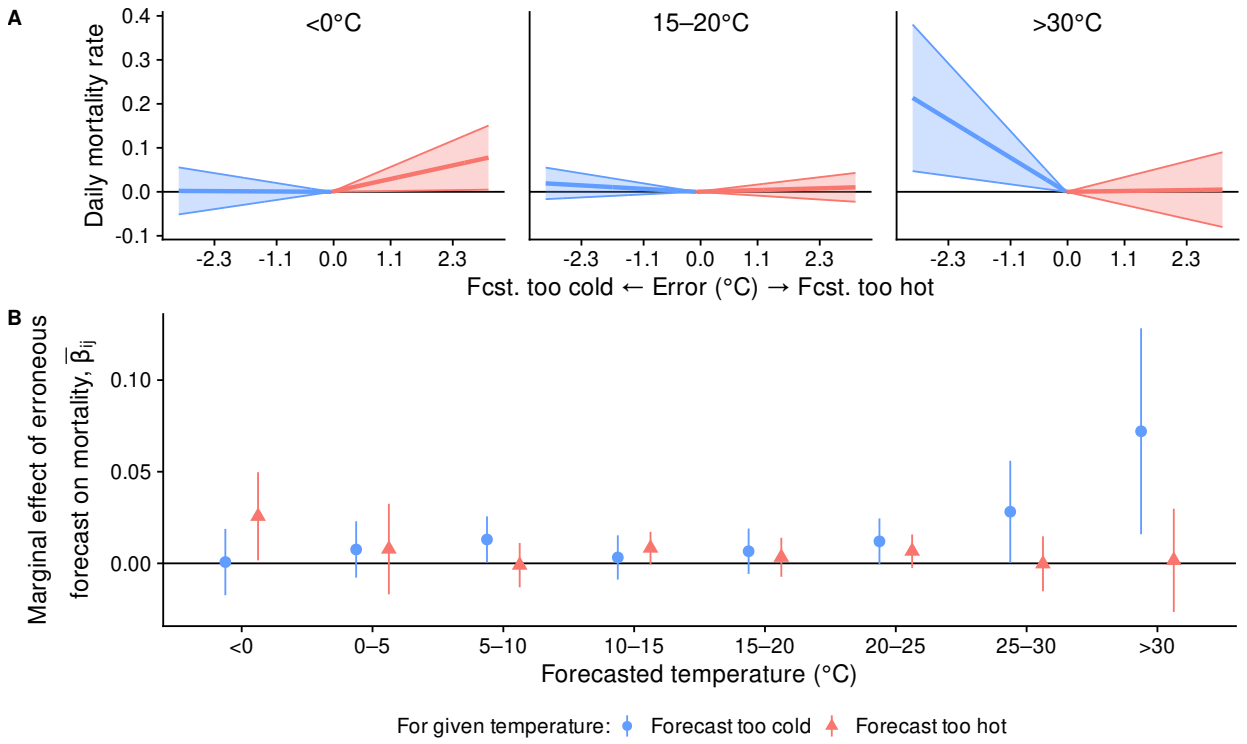
Now consider the results of estimating equation (10). Figure 2 shows 6-day cumulative effects of forecast errors on mortality based on a single estimation of equation (10). The bottom portion of the figure (panel B) shows the point estimates for $\bar{\beta}_{i,j}$, the cumulative effect of more erroneous forecasts on mortality from (11).[25] As a guide to interpretation, the top portion of the figure (panel A) shows predicted effects on mortality from the marginal effects in panel B.

The far left pair of points shows the effect of forecast errors on a day when the expected temperature is below freezing (< 0°C). The blue, circular point shows that a larger negative forecast error does not have a strong effect on mortality: a forecast saying that it will be colder than it turns out to be does not lead to more deaths. In contrast, a 1°C larger positive forecast error does increase mortality by just over 0.01 deaths per 100,000 people, representing around a 0.5% increase in mortality over the average daily rate. The leftmost plot in panel A shows the same result, but in terms of predicted mortality. Predicted mortality barely changes with a negative forecast error but rises significantly with a positive error. Accounting for both of the cold-temperature coefficients, an average forecast error on a cold day increases mortality by about 0.004 deaths per 100,000 people, or about 0.15%.[26]

[25]The forecast error effect we find is based on daily, county-wide average temperature. Thus, a 1°C error in our analysis could correspond to a substantially larger forecast error at specific locations within the county or specific moments during the day.

[26]One extra day per month of realized temperatures in this range raises mortality by about 1% relative to a mild day. In contrast, extending the realized temperature analysis from Barreca et al. (2016) to our sample period, we find that an extra hot day reduces mortality relative to a moderate day. One reason for this

Figure 2: Forecast Errors Increase Mortality



Notes: The figure shows results of a single estimation of Equation (10) on the baseline sample. The dependent variable is the daily mortality rate per 100,000 people. The reported values are 6-day (day t through day $t+5$) cumulative effects of forecast errors. The model includes covariates for lags of 5°C bins of realized temperature, lags of indicators for above median precipitation, lags of four bins in forecasted temperature, date fixed effects, and county-by-month fixed effects interacted with a linear time trend. The regression is weighted by county population. Standard errors are clustered at the CWA-level.

The other coefficients can be interpreted similarly. Overall, during periods with warm temperatures, negative forecast errors are deadly while positive forecast errors have little or no effect, on average. The reverse is true for the coldest temperatures. Reassuringly, moderate temperatures reveal little effect of forecast errors on mortality. It is intuitive that changes in behavior have smaller effects on mortality risk when that risk is already small.

Table 1 shows similar estimates to the figure, but using only four forecast bins for legibility. We will use these same bins when testing mechanisms and valuing forecast errors. The columns of the table show the effect of forecast errors on mortality for errors below the breakpoint (left column) and above the breakpoint (right column). The rows show the effects within different temperature ranges. The results are consistent with the richer set of forecasted temperature bins shown in Figure 2.

Our point estimates so far are consistent with Definition 1 (non-monotonic effects and therefore appropriate adaptation). But if we use Assumption 1 we can get more power and then we find significant convexity. The theoretical analysis showed that the convexity of mortality risk in forecast errors is critical to assessing the mortality value of forecast improvements. We assess whether the relationship between forecast errors and mortality is convex by testing whether the sum of the negative and positive forecast error marginal effects is greater than 0. The p-value is 0.35 for forecasted temperatures less than 0°C, but we reject the null hypothesis of nonconvexity at the 5% level in all other temperature ranges (with p-values of 0.035, 0.006, and 0.029 for cool, warm, and hot temperatures).[27] As a result, we conclude that adaptation is appropriate (Definition 1), that $\lim_{\epsilon \to 0} h_{AA}(T, A^*(T)) > 0$ (Appendix B.2), and that improving the accuracy of forecasts reduces expected mortality (Proposition 1).

4.3 Robustness and Sensitivity

Figure 3 compares the baseline estimates from Table 1's hot bin to estimates from a variety of different specifications. Results are largely similar across the rest of the temperature distribution and can bee seen in Section E. We separate results into changes in clustering,

unexpected result is that conventional analyses of realized temperature conflate expected and unexpected forecast shocks: writing forecasts as a function of T and assuming e is independent of T, they recover $\frac{\mathrm{d}}{\mathrm{d}T} h(T, A^*(T+e)) = h_T + \mathrm{d}h/\mathrm{d}e$ (averaged over observed forecast errors). Throughout the manuscript we find that forecasts are especially effective in hot temperatures, so that the two terms plausibly conflict. And in the quadratic analysis of Section 6, we do find that the two terms roughly offset each other (with h_T positive and quite large) in the hot bin. Complementary explanations include that: trends in air conditioning adoption analyzed by Barreca et al. (2016) have continued into our more recent sample period; forecasts have improved the most on hot days during our sample (see Figure 1b); and public health authorities encourage individuals to go to cooling centers on hot days, where they may socialize or receive additional medical attention.

[27] Estimating a quadratic specification, we find quantitatively similar results except that in the hot bin the convexity is no longer significant (see Table 3).

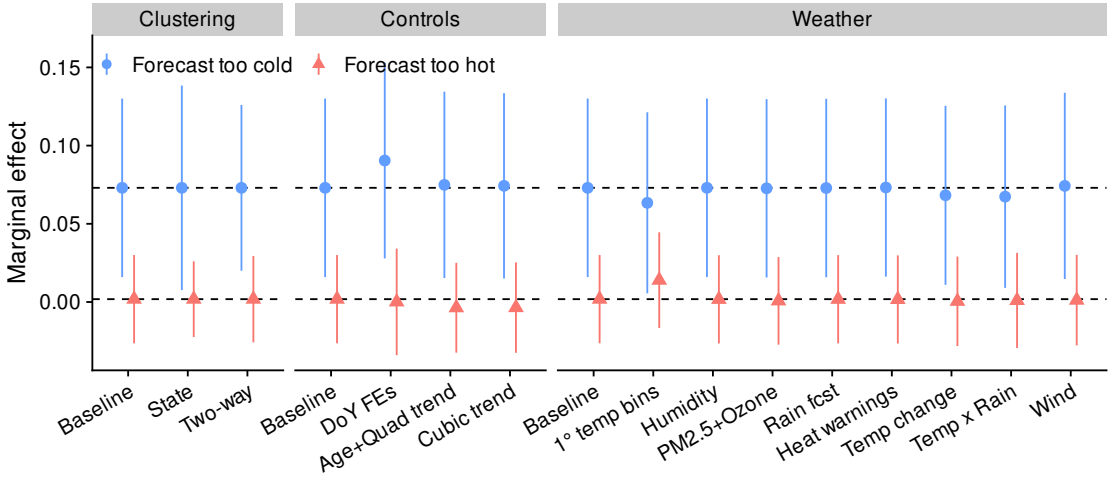
Table 1: Testing Monotonicity of Mortality-Forecast Error Relationship

	(1) Negative error *Forecast too cold*	(2) Positive error *Forecast too hot*
<0°C (cold)	-0.005	0.013**
	(0.006)	(0.005)
0–15°C (cool)	0.007*	0.004
	(0.004)	(0.004)
15–30°C (warm)	0.013***	0.002
	(0.005)	(0.003)
>30°C (hot)	0.073**	0.002
	(0.029)	(0.014)
Avg. death rate	2.245	
Clusters	130	
N	13,642,033	

Notes: The table shows results of a single estimation of Equation (10) on the baseline sample. The dependent variable is the daily mortality rate per 100,000 people. The reported values are 6-day cumulative effects of larger forecast errors. Column 1 shows negative forecast errors (below median for each expected temperature bin), and Column 2 shows positive forecast errors (above median). The model includes covariates for 6 lags of 5°C bins of realized temperature, indicators for above median precipitation, and four bins of forecasted temperature plus date fixed effects and county-by-month fixed effects interacted with a linear time trend. Weighted by county population. Standard errors, clustered at the CWA-level, are below each estimate. Significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

changes to the non-weather controls, and changes to the controls for realized weather and other forecasts. Across the full set of checks, the point estimates remain largely unchanged, so the text below focuses on some highlighted checks and limits the discussion to effects during periods with hot forecasted temperatures. Appendix E contains a full discussion of all checks and forecasted temperatures.

Figure 3: Robustness and Sensitivity Checks: Hot Forecasted Temperature



Notes: The figure shows robustness and sensitivity checks on the main results reported in Section 4 for hot expected temperatures. The other temperature ranges are shown in Section E. All models use the same functional form and lag length as the baseline results. The lines are 95% confidence intervals based on standard errors clustered at the CWA level. The labels indicate the change or addition. For comparison, "Baseline" reproduces the baseline estimate, with controls described in Section 4.1. Unless otherwise specified, added weather variables are controlled for non-parametrically using quantile bins. In all cases, 6 lags of the variables are included to match the lag length of the forecast error. Note that wind and rainfall forecasts are only available for a subset of the observations, so the sample changes.

Both temperature and mortality are highly seasonal. Including finer location-specific seasonality controls by replacing month-of-year fixed effects with day-of-year fixed effects makes the results slightly stronger. However, our base specification's more parsimonious month controls do appear to largely account for seasonal patterns.

Realized weather on day t is not in the forecaster's information set and could bias our estimates if it is correlated with forecast errors. The checks in the "Weather" section of the plot show that this bias may not be of practical concern. Controlling for humidity, pollution, or wind affects the estimate only minimally. Using finer bins of realized temperature (1°C versus 5°C in the baseline) reduces the effect of negative forecast errors but increases the effect of positive forecast errors, leaving the average effect unchanged.

In addition to realized weather, one might be concerned that correlated forecasts or forecast errors might be driving the results. Temperature is one of the most important drivers

of mortality among all weather phenomena, so temperature forecasts likely constitute the majority of the effect we find. We assess the effect of correlated rainfall forecasts by adding them to the baseline regression ("Rain fcast"), using bins for above and below median as with the realized rainfall controls. The results are unaffected.

The results also do not change when excessive heat warning indicators are added as controls to the model. The NWS issues excessive heat warnings when heat is expected to rise above dangerously high thresholds. These warnings are therefore correlated with temperature forecasts in the hot and warm bins. The robustness results show, however, that holding constant the presence or absence of an active warning does not affect the results. Our baseline specification does include the effects of forecast-driven heat warnings in the estimated total effects of forecasts, but the present results imply that forecasts drive our baseline results through channels other than their role in issuing formal warnings.

5 Mechanisms

We have seen that forecast errors matter for mortality, holding realized temperatures constant. Agents must be using forecasts to take actions that affect their mortality (see Lemma 3 in Appendix B.2). We now explore what these actions may be.

5.1 Evidence from Time Use and Electricity Consumption

To test for mechanisms, we estimate:

$$z_{ct} = \sum_{\ell=0}^{L} \left[\sum_{j=1}^{J} \mathbf{1}\{f_{c,t-\ell} \in B_j\}(\beta_{20,\ell,j} + \beta_{21,\ell,j} e_{c,t-\ell}) \right] + \sum_{\ell=0}^{L} \left[g_{1,\ell}(T_{c,t-\ell}; \xi_{21}) + g_{2,\ell}(\text{prec}_{c,t-\ell}; \xi_{22}) \right] + X_{ct}\gamma_2 + \alpha_{2,cm} + \rho_{2,t} + \varepsilon_{2,ct}. \quad (12)$$

All variables are as in equation (10) except that the dependent variable z_{ct} is here either time use or electricity consumption. We estimate linear relationships between outcomes z and forecast errors because actions may depend linearly on forecasts even as mortality depends nonlinearly on forecast errors.[28]

The time use analysis relies on daily, individual-level data from the American Time Use Survey (ATUS). These data are geocoded at the county, state, or core-based statistical area (CBSA) level, and we observe data for 100 out of the 130 County Warning Areas in the continental U.S. The electricity data are at the state-month level and are from the

[28] Appendix I reports nonlinear regressions of actions on forecast errors in Tables A6 and A4.

U.S. Energy Information Agency (EIA). We aggregate the daily, county-level forecast data to the state-month level by taking population-weighted averages.[29]

Table 2: Residential Electricity Demand and Time Use

	(1)	(2)	(3)	(4)
	Time use (minutes/day)			Log electricity
	Work	Home prod.	Leisure	demand
< 0°C × Forecast error	3.94	-2.03	-1.91	0.00029
	(4.41)	(4.00)	(7.16)	(0.00026)
0 to 15°C × Forecast error	0.53	-5.84***	5.31**	-0.00029
	(2.57)	(1.99)	(2.51)	(0.00020)
15 to 30°C × Forecast error	3.95	-1.63	-2.32	0.00014
	(2.78)	(2.45)	(2.52)	(0.00019)
> 30°C × Forecast error	12.5	27.6***	-40.1***	0.0014***
	(8.07)	(10.1)	(7.37)	(0.00033)
LHS mean	189.8	263.8	986.4	
N	144,234	144,234	144,234	7,104
Clusters	100	100	100	48

Notes: The table shows results of estimating 4 versions of Equation 12 on two different datasets. Columns 1 through 3 use daily, individual-level time use data from ATUS. The dependent variables are time uses in minutes and are indicated at the top of each column. The forecasted temperature bins are based on daily average temperature forecasts. The coefficients are 6-day cumulative effects. Column 4 uses monthly, state-level electricity demand data from EIA. The dependent variable is the log of monthly residential electricity demand. The right-hand-side forecasted temperature bins are the number of days per month forecasted to be in those ranges. The models include covariates for 5°C bins of realized temperature, quartiles in precipitation (Columns 1–3) or a quadratic in precipitation (Column 4), four bins for forecasted temperature, as well as time (date for Columns 1–3, year-month for Column 4) fixed effects, and county-by-month fixed effects interacted with a linear time trend. Columns 1-3 are weighted by county population and Column 4 by state population. Standard errors are clustered at the WFO (1–3) or state (4) level. Significance: * $p < .10$, ** $p < .05$, *** $p < .01$.

The time use regressions in Columns 1–3 of Table 2 show that 1-day-ahead forecasts cause people to change their time use patterns, particularly on hot days. For a day forecasted to be above 30°C on average, individuals respond to a 1°C higher forecast error (equivalently, a 1°C hotter forecast) by reducing their leisure time by 40 minutes and shifting their time use toward work and home production, even though we hold realized temperature fixed. The effect of forecasts is smaller on milder or colder days, but there is still about 4 to 5 minutes

[29] Appendix A provides details on each dataset and on the definition of each time use category.

of total time-shifting on those days in response to forecasts. And the pattern from hot days reverses on cold days: when we see a warmer forecast in cold weather, people spend more time on leisure activities and less time on home production. People do respond to forecasts by meaningfully changing their activities.

Column 4 shows that there is also a significant effect of forecasts on hot days' electricity use. A 1°C larger forecast error increases residential electricity demand by 0.14%. Table A5 reports estimates for all variables in the model, including realized temperature. The table shows that an extra day forecasted to be above 30°C, compared to a day between 15 and 20°C, leads to 1.2% higher electricity demand, roughly one-tenth the size of the effect of an extra hot day. Viewed in the context of the results in Columns 1–3, this additional residential electricity demand could come from the increased use of electricity by individuals staying home rather than going out to work or engaging in leisure.

5.2 Cross-Sectional Evidence

We investigate additional mechanisms by examining the correlation between our estimated county-level effect of forecasts on mortality and geographical characteristics. First, we estimate county-level versions of equation (10) to recover coefficients $\bar{\beta}^c_{11,j}$ and $\bar{\beta}^c_{12,j}$, with superscript c indicating the county.[30] Next, for each county, we calculate the average convexity of the forecast-mortality function within forecasted temperature bin: $\hat{\bar{\beta}}_{c,j} \triangleq (1/2)(\hat{\bar{\beta}}^c_{11,j} + \hat{\bar{\beta}}^c_{12,j})$ for $j \in \{1, 2, 3, 4\}$.

We use the county-level estimates of average convexity in the following regression:

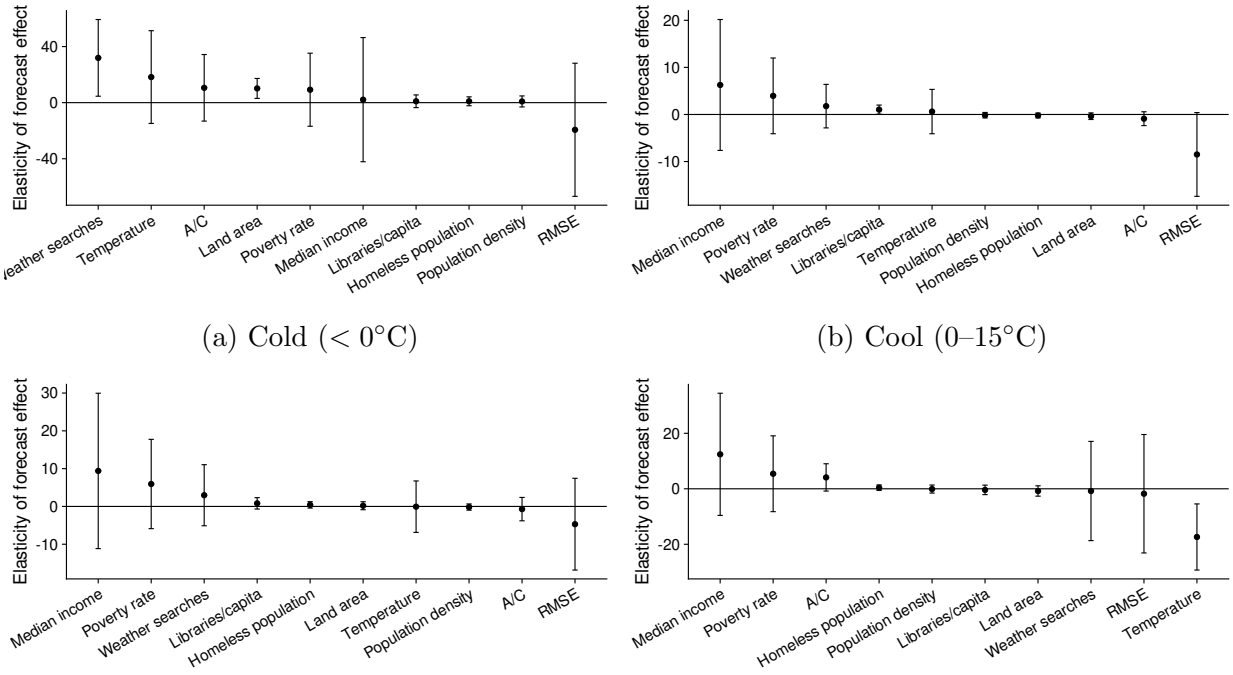
$$\hat{\bar{\beta}}_{c,j} = Z_c\theta + \sum_{i=1}^{2} \left[\Gamma_{1,i}(Lat_c)^i + \Gamma_{2,i}(Long_c)^i\right] + \alpha + \varepsilon_c. \tag{13}$$

One regression is run for each forecasted temperature bin, j. The matrix Z_c contains the variables of interest: average temperature; county-level median income and poverty rate from the Census Small Area Income and Poverty Estimates (SAIPE) program; Google Trends searches at the state level for "weather" normalized by searches for "dog"; libraries (which act as heating and cooling shelters during extreme weather) per capita from the Institute of Museum and Library Services Public Library Survey; home air conditioning adoption, derived from American Community Survey and American Housing Survey data; land area from the 2010 Census county shapefile; the count of homeless persons per state, from the

[30] These are county-level analogies of the coefficients in Equation 11. Regressions are run separately for each county. We replace county-by-month fixed effects with month fixed effects. In order to reduce the noisiness of estimates (e.g., in low-population counties that rarely experience mortality events), the estimates are partially pooled to the average value across locations using an empirical Bayes procedure. The procedure shrinks estimates toward the group mean in proportion to the standard error of the county-level estimate.

Department of Housing and Urban Development; population density, derived from SEER population data; and the RMSE of day-ahead temperature forecasts over the sample period.[31] For all variables, we take simple averages over the sample period to create cross-sectional measures that vary by county c. In addition to the depicted variables, the regression controls for quadratics in the latitude and longitude of the county centroid. The regression uses heteroskedasticity robust standard errors.

Figure 4: Correlation Between County-Level Covariates and Average Convexity



(a) Cold (< 0°C)

(b) Cool (0–15°C)

(c) Warm (15–30°C)

(d) Hot (> 30°C)

Notes: The figure shows the elements of $\hat{\theta}$ estimated from regression (13). The points are coefficient estimates and whiskers are 95% confidence intervals from heteroskedasticity-robust standard errors.

Figure 4 plots the estimated $\hat{\theta}$. Overall, the signal to noise ratio is low for these estimates, but some consistent patterns emerge. Higher median income and poverty rates are associated with a stronger forecast effect. Higher income might allow for more flexibility when responding to forecasts, but locations with higher poverty rates likely have a larger population at high risk of mortality from temperature and a greater need for forecast-driven public health interventions. Income effects are largest for hotter forecasted temperatures, whereas poverty effects are uniform across the temperature bins, lending support to the possibility that adaptation activities at hotter temperatures are costly. Higher RMSE (less accurate

[31] See Appendix A for details on the datasets and Appendix A.2 for a description of AC estimation procedure.

forecasts on average) reduces forecast effectiveness across all bins. As discussed further in Appendix C, this result is consistent with individuals finding forecasts less useful when they are typically noisy.[32]

Higher air conditioner penetration is also associated with a stronger forecast effect in both extreme bins, which suggests that air conditioning does not substitute for forecasts but is, if anything, a complement (in line with the electricity demand evidence from Section 5).

Average temperature has an asymmetric effect. In the cold bin, higher temperatures raise the forecast effect while in the hot bin it lowers it, potentially because extreme temperatures are more surprising in these locations, making forecasts more helpful.

5.3 Survey Evidence

To complement the revealed preference evidence of the impact of forecast on actions, we also collect stated preference evidence using a survey of 128 University of Arizona students.[33] Although a sample of convenience, this group can still be considered an informative case as respondents should be relatively unlikely to act on forecast information: they are of an age group that has low sensitivity to extreme temperature, and they took the survey during a season of relatively mild temperatures.[34]

Overall, respondents cared about the weather forecast: 90% of respondents replied that they cared "a little bit" or "quite a lot" about what the weather will be tomorrow, as opposed to not "at all" (10%). In addition, 82% of respondents responded that they look at the weather forecast at least once a week and 52% reported looking at the forecast at least once a day (compared to 60% for a recent representative sample in the U.S. (Orth, 2023)). Respondents were further asked to "think back to a time during the summer when the temperature was hot," and asked, "If the temperature was forecast to be especially hot the next day, would you plan to do anything differently because of the forecasted hot weather?" 69% of respondents replied "Yes". When asked, "What would you do differently?", open-ended responses mentioned avoiding outdoor activity, drinking water, changing wardrobe, or finding a cool activity. In addition, of the 89% of respondents who recalled "a time when there was an excessive heat warning where you lived," 61% responded that they did something differently because of the warning. A majority of this subset of respondents recalled stay-

[32] In concurrent work, Song (2022) examines effects of RMSE on labor supply in China. Our estimated effects of awareness of forecast accuracy are consistent with Song's results, and our identification is similar in being cross-sectional.

[33] The survey (STUDY00001720) was reviewed and approved by the University of Arizona Human Subjects Protection Program and administered through the Arizona Policy Lab. Student were incentivized with one point of extra credit in a participating class of their choice for taking the survey. The survey was run from September 2022 to May 2023.

[34] Average day-ahead high temperature forecast during the study period was 77.7°F, with a minimum of 52°F and a maximum of 98°F.

ing inside as a response to the warning, and others noted modifying the times of outdoor activities (dog walking, yard work, etc.) to cooler parts of the day. Some respondents also recalled behavioral modifications like drinking more water or wearing weather-appropriate clothing. Finally, 92% of respondents replied "Yes" to the question "Do you think excessive heat warnings give the public information that helps to save lives?" While these responses could suffer from well-known survey biases (e.g., recall bias, social desirability bias), they do corroborate our revealed preference findings.

6 Effect of Forecast Accuracy on Mortality

Having seen that forecasts matter for mortality, and having seen that several lines of evidence suggest that people do use forecasts in ways consistent with our estimated effects on mortality, we now use the theoretical analysis to quantify the effects of improved forecast accuracy. In particular, we estimate the average mortality benefit following Proposition 1 and willingness to pay following Proposition 2.

6.1 Estimating Equation

Each calculation requires second-order effects of forecast errors and realized temperature on mortality risk. We therefore estimate:

$$\begin{aligned} y_{ct} = &\sum_{\ell=0}^{L}\sum_{j=1}^{J} \mathbf{1}\{f_{c,t-\ell} \in B_j\}\Big(\beta_{30,\ell,j} + \beta_{31,\ell,j} e_{c,t-\ell} + \beta_{32,\ell,j} e_{c,t-\ell}^2 \\ &\qquad + \beta_{33,\ell,j} T_{c,t-\ell} + \beta_{34,\ell,j} T_{c,t-\ell}^2 + \beta_{35,\ell,j} e_{c,t-\ell} T_{c,t-\ell}\Big) \qquad (14) \\ &+ \sum_{\ell=0}^{L} g_{2,\ell}(\text{prec}_{c,t-\ell}; \xi_{32}) + X_{ct}\gamma_2 + \alpha_{3,cm} + \rho_{3,t} + \varepsilon_{3,ct}. \end{aligned}$$

All variables are as in equation (10). Here we include an interacted quadratic in realized temperature and forecast errors. The mortality benefit will depend on the $\beta_{32,\ell,j}$, and willingness to pay will depend on the $\beta_{34,\ell,j}$ and the $\beta_{35,\ell,j}$.

6.2 Results

We saw in Section 4 that mortality risk is convex in forecast errors. Therefore, from Proposition 1, increasing the accuracy of forecasts will save lives when Assumption 1 holds. We quantify this effect for two counterfactual changes in forecast accuracy. Our calculations assume: (i) that the estimated changes in mortality are persistent, and (ii) that the estimated response to forecasts is invariant to the postulated changes in forecast accuracy. We validate

these assumptions in Appendix C, and we also test robustness of the results to changes in functional form.[35]

The top panel of Table 3 reports data-dependent parameters. The top row shows that cold and hot days are far less common than cool and warm days, and the second row shows that forecasts are more accurate on hotter days. Our mortality benefit and net value calculations will rely on an approximation in which Assumption 1 holds with ϵ small, meaning that the variance of forecast errors is small and the VSL is large relative to marginal costs of adaptation. The second row shows that forecasts are in general fairly accurate, and following EPA (2021), we use a VSL of \$9.5 million (in 2020 dollars) that surely exceeds marginal adaptation costs by a large margin.

The middle panel considers the mortality benefit of improved forecast accuracy. Its first row shows that, as in Section 4, mortality risk is convex in forecast errors, especially on hot days. Its second row is the marginal effect of the standard deviation of forecast errors on annual, nationwide mortality.[36] Following Proposition 1, this row reports the convexity multiplied by the number of days per year of weather in that bin, the population of the U.S. in hundred thousands, and the derivative of variance with respect to standard deviation.[37]

More accurate forecasts save lives: reducing the standard deviation of forecasts errors by 1°C would save 5,400 lives per year. Across the temperature distribution, the majority of this value comes from warm and cool days because these temperatures are so much more common than extreme heat or cold.

The third row of the middle panel of Table 3 calculates lives saved if the standard deviation of forecast errors were halved. The top panel shows that forecasts of hot temperatures have the smallest standard deviation and forecasts of cold temperatures have the largest standard deviation. The difference in the spread of errors is likely due to underlying meteorology: warm air masses are more stable and easier to forecast than colder, more volatile air masses. Cold and cool days' contribution to the total mortality benefit increases from the second to the third rows of the middle panel because forecasts on these days have more room for improvement.

Thus far we have shown that improving forecasts saves lives. The monetized value of lives saved is \$51 billion from a 1°C more accurate forecast and \$21 billion from a 50% more accurate forecast. However, these calculations are not agents' own value for more accurate forecasts, which nets out their costs of acting on forecasts and accounts for the terms highlighted in Proposition 2.

[35] And the results on lives saved are similar whether we use a quadratic specification in Equation 14 or a more flexible two-line estimator (see Section C). The interacted quadratic specification is conservative and allows for estimation of WTP.

[36] Appendix F contains details on the counterfactual calculations.

[37] We use 310 million for our sample period population and a mean solar year as our year.

Table 3: Annual Mortality Benefit and Net Value From Forecast Accuracy Improvements

	Forecasted Temperature				
	< 0°C Cold	0–15°C Cool	15–30°C Warm	> 30°C Hot	All Temps
Parameters					
Frequency $n(T)$ (days/year)	35	136	188	6	365
Current Error Std Dev $\sqrt{Var[T\|f]}$ (°C)	1.40	1.23	1.03	0.91	1.15
Mortality Benefit					
Error Convexity h_{ee}	0.0015	0.0033	0.0055	0.0078	0.0044
	(0.0024)	(0.0017)	(0.0022)	(0.0185)	(0.0012)
Lives Saved from Smaller Std Dev:					
Marginal Improvement	222	1,713	3,336	136	5,406
	(369)	(867)	(1,310)	(320)	(1,526)
50% Improvement	116	790	1,291	46	2,243
	(193)	(400)	(507)	(109)	(633)
Net Value					
Temperature Convexity h_{TT}	-0.00038	-0.00012	0.00102	0.00480	0.00053
	(0.00019)	(0.00025)	(0.00025)	(0.00297)	(0.00011)
Error-Temperature Interaction h_{eT}	-0.00040	-0.00042	0.00075	-0.00484	0.00011
	(0.00054)	(0.00037)	(0.00039)	(0.00607)	(0.00038)
WTP for Smaller Std Dev ($billion):					
Marginal Improvement	-1.1	-2.6	10.1	0.0	6.3
	(0.3)	(1.2)	(1.4)	(0.5)	(1.3)
50% Improvement	-0.6	-1.2	3.9	0.0	2.1
	(0.1)	(0.6)	(0.6)	(0.2)	(0.4)

Notes: The table shows the mortality benefits and willingness to pay for more accurate weather forecasts, both broken down by forecasted temperature bins and across all temperatures. The first section shows the frequency of each temperature level and the associated standard deviation of forecast errors. The second section shows the convexity of the relationship between mortality and forecast errors estimated by Equation (14), along with the expected annual lives saved from forecast error reductions in two counterfactual scenarios. It measures the mortality rate increase from a marginal change in mean absolute forecast error. The third panel shows the convexity of mortality with respect to temperature and the cross-partial of forecast error and temperature, generated from the same estimation of Equation (14). From these estimates, net value of forecast improvements is calculated under the same two counterfactuals.

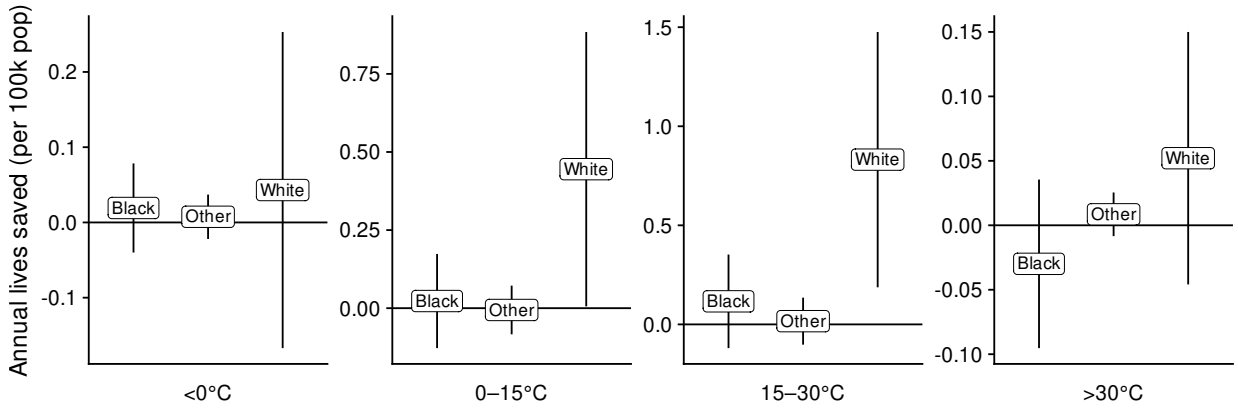
The bottom panel of Table 3 values more accurate forecasts as agents' willingness to pay for them, which accounts for their costs of acting on forecasts. Proposition 2 showed that this value depends on the convexity of mortality in temperature (first row) and on how responses to accurate forecasts mitigate that convexity (second row). Mortality risk is convex in temperature on warm days and especially convex on hot days. In line with the theory, the error-temperature interaction typically has a negative (albeit noisy) point estimate. Agents use forecasts to smooth mortality risk on hot days to such a degree that they receive little net value from having more accurate forecasts on hot days (final two rows). Agents' responsiveness to hot forecasts is consistent with foregoing results showing that forecast errors are especially important on hot days and that time use and electricity use are especially responsive to forecasts on hot days. This smoothing effect vanishes on warm days, reflecting either that actions are less responsive to forecasts on warm days (so that $A^{*\prime}$ is small) or that actions' effects on mortality risk are less temperature-dependent on warm days (so that h_{AT} is small). As a result, people would pay \$10 billion for a 1°C reduction in the standard deviation of forecast errors on warm days and \$4 billion for a 50% reduction.

6.3 Heterogeneity Analysis

To assess who is benefiting from forecasts, we estimate versions of equation (14) where the outcome variable is replaced with mortality for different demographic groups, as reported on the death certificates. Figure 5 reports these results for different race groups, and Figures A6b, A6a, and A7 in Appendix G report results by age, sex, and cause of death. The location-specific heterogeneity in Section 5 also provides evidence on variation in lives saved, while regional heterogeneity is presented in Figure A9. The main takeaways from this analysis are: (1) cause of death generally follows the same pattern as seen in the effect of realized temperature on mortality (cardiovascular disease being particularly strongly associated with forecast errors), but forecasts are strongly associated with accidents while realized temperatures are not; (2) adaptation is "protective" for women in cold but otherwise similar for men and women; (3) individuals 65 or older respond most strongly to forecasts and especially so in colder weather; (4) colder parts of the country benefit from forecasts in hot weather and hotter parts of the country benefit from forecasts in cold weather; and (5) mortality for white people responds more strongly to forecasts.

Figure 5 shows that this last result holds across the temperature distribution. The heterogeneity could be due to differences in preferences, information, credibility, or constraints on adaptive behavior. Prior work suggests that the first possibility is unlikely given that realized temperature is a relatively more important source of mortality for people of color than for white Americans (Basu, 2009). The second and third possibilities would imply

Figure 5: Heterogeneity in Lives Saved by Race of the Deceased



Notes: The figure shows the annual lives saved (per 100,000 people) from a 1°C reduction in the standard deviation of forecast errors. Estimates for each demographic group come from a separate model fit using Equation (14) on the baseline data and translated into annual lives saved by multiplying the convexity by the days per year in each forecasted temperature bin. The lines are 95% confidence intervals based on standard errors clustered at the CWA level.

benefits from a policy to improve communication and outreach, whereas the fourth would imply benefits from a policy to relax constraints on adaptation. Future work should explore which mechanisms are especially important.

6.4 Projected Benefits of Forecast Improvements Under Climate Change

As the climate in the U.S. warms, both warm and hot days will become more frequent. From Table 3, these days have the greatest benefit in terms of lives saved and warm days have the greatest monetized net benefits. Therefore, as the climate warms, forecasts will become more valuable, ceteris paribus.[38] Table 4 shows the projected change in forecast value from the changing distribution of temperatures in the U.S. over the coming century under SSP2-4.5, a so-called "middle of the road" scenario in which emissions begin to fall midcentury and end-of-century warming is 2.5–3°C.[39]

The top panel of the table shows days per year in each temperature bin, the lives saved per year, and the net value (WTP) from a 50% forecast improvement in 2100. Even under

[38]In order to isolate the role of changing weather, our climate calculations hold the population's level, demographic structure, geographic distribution, and adaptation options constant over time. Our calculations also assume that people's use of forecasts does not change as they experience climate change (see Guido et al., 2021) and that the short-run forecastability of weather does not change with climate change (see Scher and Messori, 2019).

[39]The specific model runs used in the projections come from the CMIP6 ScenarioMIP runs of the NOAA Geophysical Fluid Dynamics Laboratory model (Guo et al., 2018). Days per year in each temperature bin are debiased with respect to the observed frequency of days from our data by matching the projections to observations during the period of overlap (2015–2017).

Table 4: Projected Forecast Value Under Climate Change

	Forecasted Temperature				All
	< 0	0–15	15–30	> 30	temps
Flow value in 2100, 50% forecast improvement					
Days/year	3	140	203	19	365
Lives saved	10	813	1396	143	2361
WTP ($ billion)	0.0	-1.3	4.2	0.0	2.9
Change from 2015	-92%	3%	8%	208%	5%
Present value of 50% forecast improvement (2025-2100)					
Lives saved (undiscounted)	3,015	61,634	102,551	6,942	174,142
Effect of climate change (lives)	-5,723	2,437	5,721	3,462	5,897
WTP ($ billion, $r = 1.7\%$)	-10.6	-54.7	177.6	-0.2	112.1
Effect of climate change ($ billion)	15.0	-2.2	8.8	-0.1	21.5

Notes: The table shows projections based on changes in realized temperature according to the CMIP6 SSP2-4.5 climate scenario. Lives saved and WTP are based on the 50% forecast improvement counterfactual given in Table 3.

this middle-of-the-road scenario, the number of cold days falls almost to zero, whereas the number of hot days more than doubles. Altogether, climate change increases lives saved and annual willingness to pay by 5%.

The bottom panel of the table shows the present value (undiscounted for lives saved, discounted at a rate of 1.7% for WTP) for a hypothetical policy that reduces the standard deviation of forecast errors by 50% in 2025–2100. Current willingness to pay for this policy is $112 billion (third row), which is $22 billion larger than in the absence of further climate change (fourth row). Therefore, a policy that obtained 50% greater forecast accuracy at a present cost of $112 billion or less would generate net benefits, even before considering non-mortality benefits of forecasts. Figure 1b showed that forecast root-mean squared error fell by 27% over the sample period. A 50% improvement could be achieved in 22 years if the recent pace of improvement were maintained. If that improvement could be obtained by redirecting all of the roughly $1 billion NWS annual budget toward this task for those 22 years, then that investment would have a present cost of $16 billion using the same discounting procedure as in Table 4. Our estimated willingness to pay for mortality benefits is large enough to justify even that extreme budget allocation.[40]

[40]Over the last decade, NOAA had a total budget of $5–6 billion, with around $1 billion allocated to the National Weather Service (CSR, 2022). The NWS is not solely responsible for forecast-related spending at the federal level, nor is the entire NWS budget dedicated to forecast improvements. For example, in 1999 the U.S. government reported spending $2.2 billion on producing and disseminating forecasts and $0.5 billion on research to improve them, with the private sector spending another $1 billion broadcasting the forecasts (Hooke and Pielke Jr., 2000).

7 Conclusion

Routine weather forecasts are a widely used, sophisticated prediction product that most people interact with on a daily basis. Despite the ubiquity of weather forecasts, the number of people who rely on them, and the global effort involved in their production, surprisingly little is known about their economic value. This paper provides the first revealed preference estimates of the value of daily weather forecasts.

We show that whether improving forecasts' accuracy reduces mortality is theoretically ambiguous, depending on the convexity of mortality risk in forecast errors and thereby on the form of adaptation undertaken. Using the universe of mortality events and weather forecasts for a twelve-year period in the U.S., we show that forecasts are effective at helping people avoid mortality, especially during extreme heat. Across the full temperature distribution, NWS forecasts save thousands of lives per year compared to less accurate alternatives and agents would pay billions of dollars for more accurate forecasts. These values are large relative to the $1 billion annual budget of the U.S. National Weather Service, suggesting that investing in improved forecasts could generate attractive social returns.

Our analysis identifies the value of more accurate forecasts through the effects of idiosyncratic forecast errors on mortality. When we value improved accuracy, we implicitly hold agents' responses to any particular forecast fixed. This approach is sensible for marginal changes in forecast quality, but nonmarginal changes could eventually change the way people use forecasts. Our results suggest that people do in fact act on forecasts more in counties where forecasts tend to be of higher quality. Our estimates may therefore represent a lower bound on the mortality value of increased accuracy. Future work should seek quasirandom persistent changes in forecast quality that can determine whether people use forecasts differently once they understand that forecasts have improved.

We show that routine weather forecasts will be an important facilitator of adaptation to climate change. Our projections assume constant forecast quality for days of a given temperature. In practice, forecast quality might improve or degrade as the climate changes, and the value of forecasts on different days could change as adaptation technology and behaviors change. Further, forecast quality is inequitably distributed around the world. International adaptation funds should consider investing in weather forecasting as one way to enhance resilience to climate change.

References

Adler, R. S. and R. D. Pittle (1983). Cajolery or command: Are education campaigns an adequate substitute for regulation? *Yale Journal on Regulation 1*(2), 159–194.

Anand, V. (2022). The value of forecast improvements: Evidence from advisory lead times and vehicle crashes. *Available at SSRN 4206910*.

Anderson, B. G. and M. L. Bell (2009). Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Epidemiology 20*(2), 205–213.

Andersson, H. and N. Treich (2011). The value of a statistical life. In A. de Palma, E. Quinet, and R. Vickerman (Eds.), *Handbook of Transport Economics*. Elgar.

Bakkensen, L. (2016). Public versus private risk information: Evidence from U.S. tornadoes. Technical report, University of Arizona.

Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2016). Adapting to climate change: The remarkable decline in the U.S. temperature-mortality relationship over the 20th century. *Journal of Political Economy 124*(1), 105–159.

Barwick, P. J., S. Li, L. Lin, and E. Zou (2019). From fog to smog: The value of pollution information. *NBER Working Paper*.

Basu, R. (2009). High ambient temperature and mortality: A review of epidemiologic studies from 2001 to 2008. *Environmental Health 8*(1), 1–13.

Bauer, P., A. Thorpe, and G. Brunet (2015). The quiet revolution of numerical weather prediction. *Nature 525*(7567), 47–55.

Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter (2018). 100 years of progress in forecasting and NWP applications. *Meteorological Monographs 59*(1), 13.1–13.67.

Berger, M. C., G. C. Blomquist, D. Kenkel, and G. S. Tolley (1987). Valuing changes in health risks: A comparison of alternative measures. *Southern Economic Journal 53*(4), 967–984.

Bickel, J. E. and S. D. Kim (2008). Verification of the Weather Channel probability of precipitation forecasts. *Monthly Weather Review 136*(12), 4867–4881.

Brooks, H. E., A. Witt, and M. D. Eilts (1997). Verification of public weather forecasts available via the media. *Bulletin of the American Meteorological Society 78*(10), 2167–2178.

Cameron, A. C. and D. L. Miller (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Carleton, T., A. Jina, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. Nath, 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.

Chapman, R. E. (1992). Benefit-cost analysis for the modernisation and associated restructuring of the National Weather Service. Technical Report 4867, National Institute of Standards and Technology.

CIESIN (2017). U.S. Census grids 2010. https://sedac.ciesin.columbia.edu/data/collection/gpw-v4. Accessed: 2018-08-30.

Craft, E. D. (1998). The value of weather information services for nineteenth-century Great Lakes shipping. *American Economic Review 88*(5), 1059–1076.

CSR (2022). National Oceanic and Atmospheric Administration (NOAA) FY2022 budget request and appropriations. Technical report, Congressional Research Service.

Deschênes, O. and E. Moretti (2009). Extreme weather events, mortality, and migration. *The Review of Economics and Statistics 91*(4), 659–681.

Dorfleitner, G. and M. Wimmer (2010, June). The pricing of temperature futures at the Chicago Mercantile Exchange. *Journal of Banking & Finance 34*(6), 1360–1370.

Drèze, J. (1962). L'utilité sociale d'une vie humaine. *Revue Fran caise De Recherche Opérationnelle 23*, 93.

Duflo, E. and E. Saez (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics 118*(3), 815–842.

EPA (2021). Mortality risk valuation. https://www.epa.gov/environmental-economics/mortality-risk-valuation. Accessed: 2021-12-10.

Fishback, P. V., W. Troesken, T. Kollmann, M. Haines, P. W. Rhode, and M. Thomasson (2011, May). Information and the impact of climate and weather on mortality rates during the Great Depression. In G. D. Libecap and R. H. Steckel (Eds.), *The Economics of Climate Change: Adaptations Past and Present*, pp. 131–167. University of Chicago Press.

Freebairn, J. W. and J. W. Zillman (2002). Economic benefits of meteorological services. *Meteorological Applications 9*(1), 33–44.

Gasparrini, A., Y. Guo, M. Hashizume, E. Lavigne, A. Zanobetti, J. Schwartz, A. Tobias, S. Tong, J. Rocklöv, B. Forsberg, et al. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet 386*(9991), 369–375.

Glahn, H. R. and D. P. Ruth (2003). The new digital forecast database of the National Weather Service. *Bulletin of the American Meteorological Society 84*(2), 195–202.

Guido, Z., S. Lopus, K. Waldman, C. Hannah, A. Zimmer, N. Krell, C. Knudson, L. Estes, K. Caylor, and T. Evans (2021). Perceived links between climate change and weather forecast accuracy: New barriers to tools for agricultural decision-making. *Climatic Change 168*(1), 9.

Guo, H., J. G. John, C. Blanton, C. McHugh, S. Nikonov, A. Radhakrishnan, K. Rand, N. T. Zadeh, V. Balaji, J. Durachta, C. Dupuis, R. Menzel, T. Robinson, S. Underwood, H. Vahlenkamp, M. Bushuk, K. A. Dunne, R. Dussin, P. P. Gauthier, P. Ginoux, S. M. Griffies, R. Hallberg, M. Harrison, W. Hurlin, P. Lin, S. Malyshev, V. Naik, F. Paulot, D. J. Paynter, J. Ploshay, B. G. Reichl, D. M. Schwarzkopf, C. J. Seman, A. Shao, L. Silvers, B. Wyman, X. Yan, Y. Zeng, A. Adcroft, J. P. Dunne, I. M. Held, J. P. Krasting, L. W. Horowitz, P. Milly, E. Shevliakova, M. Winton, M. Zhao, and R. Zhang (2018). NOAA-GFDL GFDL-CM4 model output.

Guo, Y., A. Gasparrini, B. Armstrong, S. Li, B. Tawatsupa, A. Tobias, E. Lavigne, M. d. S. Z. S. Coelho, M. Leone, X. Pan, et al. (2014). Global variation in the effects of ambient temperature on mortality: A systematic evaluation. *Epidemiology 25*(6), 781.

Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of Economic Literature 61*(1), 3–40.

Haas, J. E. and R. B. Rinkle (1979). Report of a study to estimate economic and convenience benefits of improved local weather forecasts. Technical Report PROFS-1, NOAA Environmental Research Laboratory, Boulder, CO.

Hastings, J. S. and J. M. Weinstein (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly Journal of Economics 123*(4), 1373–1414.

Hooke, W. H. and R. A. Pielke Jr. (2000). Short-term weather prediction: An orchestra in search of a conductor. In D. Sarewitz, R. A. Pielke Jr., and R. Byerly (Eds.), *Prediction: Science, Decision Making and the Future of Nature*, pp. 61–83. Island Press.

Jones-Lee, M. (1974, July). The value of changes in the probability of death or injury. *Journal of Political Economy 82*(4), 835–849.

Katz, R. W. and J. K. Lazo (2011). Economic value of weather and climate forecasts. In M. P. Clements and D. F. Hendry (Eds.), *The Oxford Handbook of Economic Forecasting*. Oxford: Oxford University Press.

Kostyshak, S. (2017). Non-parametric testing of U-shaped relationships.

Kruttli, M., B. Roth Tran, and S. W. Watugala (2019). Pricing Poseidon: Extreme weather uncertainty and firm return dynamics. *Finance and Economics Discussion Series* (2019-054).

Lave, L. B. (1963). The value of better weather information to the raisin industry. *Econometrica 31*(1/2), 151–164.

Lazo, J. K., R. E. Morss, and J. L. Demuth (2009). 300 billion served: Sources, perceptions, uses, and values of weather forecasts. *Bulletin of the American Meteorological Society 90*(6), 785–798.

Letson, D., D. S. Sutter, and J. K. Lazo (2007, August). Economic value of hurricane forecasts: An overview and research needs. *Natural Hazards Review 8*(3), 78–86.

Meza, F. J., J. W. Hansen, and D. Osgood (2008). Economic value of seasonal climate forecasts for agriculture: Review of ex-ante assessments and recommendations for future research. *Journal of Applied Meteorology and Climatology 47*(5), 1269–1286.

Miller, B. M. (2018). The not-so-marginal value of weather warning systems. *Weather, Climate, and Society 10*(1), 89–101.

Millner, A. and D. Heyen (2021). Prediction: The long and the short of it. *American Economic Journal: Microeconomics 13*(1), 374–398.

Molina, R. and I. Rudik (2022). The social value of predicting hurricanes. *Working Paper*.

Morss, R. E., J. K. Lazo, B. G. Brown, H. E. Brooks, P. T. Ganderton, and B. N. Mills (2008). Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. *Bulletin of the American Meteorological Society 89*(3), 335–346.

Myrick, D. T. and J. D. Horel (2006). Verification of surface temperature forecasts from the National Digital Forecast Database over the Western United States. *Weather and Forecasting 21*(5), 869–892.

National Research Council (1998). *The Atmospheric Sciences: Entering the Twenty-First Century*. Washington, DC: The National Academies Press.

National Research Council (2010). *When Weather Matters: Science and Services to Meet Critical Societal Needs*. Washington, DC: The National Academies Press.

Neidell, M. (2009). Information, avoidance behavior, and health the effect of ozone on asthma hospitalizations. *Journal of Human Resources 44*(2), 450–478.

NOAA (2021). NOAA Blue Book FY2022. Technical report, National Oceanic and Atmospheric Administration.

Orth, T. (2023). How often and where Americans get information on the weather. https://today.yougov.com/topics/health/articles-reports/2023/05/08/how-and-where-americans-get-information-weather. Accessed: 2023-05-08.

Pielke, R. and R. E. Carbone (2002, March). Weather impacts, forecasts, and policy: An integrated perspective. *Bulletin of the American Meteorological Society 83*(3), 393–403.

PRISM Climate Group (2004). http://prism.oregonstate.edu.

Raymond, C. and S. Taylor (2021). "Tell all the truth, but tell it slant": Documenting media bias. *Journal of Economic Behavior & Organization 184*, 670–691.

Richardson, D. S. (2000). Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society 126*(563), 649–667.

Rogers, D. P. and V. V. Tsirkunov (2013, October). *Weather and Climate Resilience: Effective Preparedness Through National Meteorological and Hydrological Services*. Washington, DC: World Bank Publications.

Roll, R. (1984). Orange juice and weather. *The American Economic Review 74*(5), 861–880.

Rosenzweig, M. R. and C. R. Udry (2019). Assessing the benefits of long-run weather forecasting for the rural poor: Farmer investments and worker migration in a dynamic equilibrium model. *NBER Working Paper*.

Scher, S. and G. Messori (2019). How global warming changes the difficulty of synoptic weather forecasting. *Geophysical Research Letters 46*(5), 2931–2939.

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

Schlenker, W. and M. J. Roberts (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.

Schlenker, W. and C. A. Taylor (2021). Market expectations of a warming climate. *Journal of Financial Economics 142*(2), 627–640.

Shogren, J. F. and T. D. Crocker (1991). Risk, self-protection, and ex ante economic value. *Journal of Environmental Economics and Management 20*(1), 1–15.

Shrader, J. (2023). Improving climate damage estimates by accounting for adaptation. *Working Paper*.

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science 1*(4), 538–555.

Song, Y. (2022). The value of weather forecasts: Labor responses to accurate and inaccurate temperature forecasts in China. Technical report.

Stewart, T. R. (1997). Forecast value: Descriptive decision studies. In A. H. Murphy and R. W. Katz (Eds.), *Economic Value of Weather and Climate Forecasts*, pp. 147–182. Cambridge: Cambridge University Press.

Stratus Consulting Inc. (2002). Economic value of current improved weather forecasts in the U.S. household sector. Technical report, Stratus Consulting Inc.

Toth, Z. and R. Buizza (2019). Weather forecasting: What sets the forecast skill horizon? In *Sub-Seasonal to Seasonal Prediction*, pp. 17–45. Elsevier.

Ver Ploeg, M. and E. Perrin (2004). *Eliminating Health Disparities: Measurement and Data Needs*. National Academies Press.

Viscusi, W. K. (1989). The political economy of risk communication policies for food and alcoholic beverages. In J. F. Shogren (Ed.), *The Political Economy of Government Regulation*, Topics in Regulatory Economics and Policy Series, pp. 83–129. Boston, MA: Springer US.

Viscusi, W. K. (1993). The value of risks to life and health. *Journal of Economic Literature 31*(4), 1912–1946.

Wang, Z. and J. Zhang (2022). The value of information disclosure: Evidence from mask consumption in China.

Weinberger, K. R., A. Zanobetti, J. Schwartz, and G. A. Wellenius (2018). Effectiveness of National Weather Service heat alerts in preventing mortality in 20 US cities. *Environment International 116*, 30–38.

Wilks, D. S. (1997). Forecast value: Prescriptive decision studies. In A. H. Murphy and R. W. Katz (Eds.), *Economic Value of Weather and Climate Forecasts*, pp. 109–146. Cambridge: Cambridge University Press.

WMO, W. Bank, GFDRR, and USAID (2015). Valuing weather and climate: Economic assessment of meteorological and hydrological services. Technical Report 1153, World Meteorological Organization, World Bank, Global Facility for Disaster Reduction and Recovery, United States Agency for International Development, Geneva, Switzerland.

Yegbemey, R. N., G. Bensch, and C. Vance (2023). Weather information and agricultural outcomes: Evidence from a pilot field experiment in Benin. *World Development 167*, 106178.

Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel (2019). What is the predictability limit of midlatitude weather? *Journal of the Atmospheric Sciences 76*(4), 1077–1091.

Appendix for Fatal Errors: The Mortality Value of Accurate Weather Forecasts

A Data Processing Details

A.1 Primary Estimation Data: Mortality, Weather, and Forecasts

The raw mortality data from the CDC NCHS MCOD files report the day and county of each vital event. All events in a county on a day are added together to generate the county-level number of daily deaths. The deaths are translated into a death rate by dividing by annual county population, as described in Section 3.2.

The day-county structure of the mortality data motivates our processing choices for the PRISM weather data and NDFD forecast data. Both datasets originally provide daily observations on a consistent, high-resolution spatial grid across the U.S. For the forecast data, there are multiple potential observations per day. Forecast models are run and results are reported multiple times per day. For the minimum and maximum temperature forecasts we focus on, the major model runs occur at 12UTC and 00UTC. Based on feedback from National Weather Service meteorologists, we examine the 12UTC forecast, which is available in the early morning for locations in the U.S. and typically forms the basis of the morning forecast on local news. We aggregate the spatial grid to the county level using the following procedure:

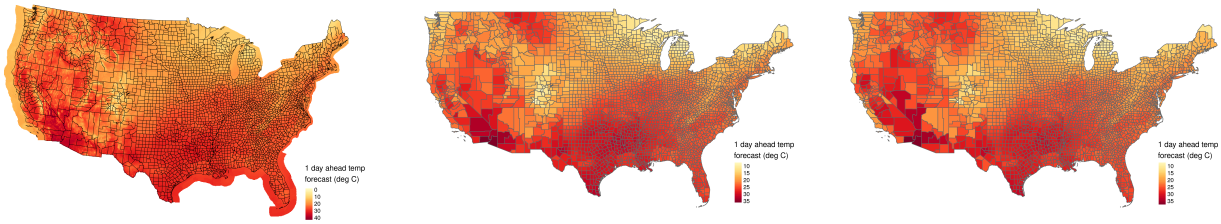
First, for each county, we find the weather and forecast grid points that fall inside the geographic boundary of the county, using 2010 county TIGER/Line shapefile from the Census. Given the high resolution of the underlying datasets (4×4 km for PRISM and either 5×5 or 2.5×2.5 km for NDFD), all counties in our sample contain multiple grid points.

Second, we assign a weight to each grid point based on 2010 population grids from CIESIN (2017). The CIESIN grids are at a roughly 1km resolution, which is higher than either the weather or forecast grid resolutions. Therefore, we use bilinear resampling to reproject the the population grid to match that of the weather and forecast grids.

Third, we calculate population-weighted average values for each weather and forecast observation within the county. The end result is a daily, population-weighted spatial average of the maximum temperature, minimum temperature, total precipitation, dewpoint temperature, maximum temperature forecast for 1 to 6 days ahead, and minimum temperature forecast for 1 to 6 days ahead (the NWS issues forecasts out to 7 days, but given our choice of the 12UTC forecast, the 7-day-ahead minimum temperature forecast is not available). Comparison of an example gridded forecast data and the corresponding county-level data is shown in Figure A1.

Fourth, we correct errors in the forecast data. The NDFD data undergo error checking (such checks are the responsibility of local Weather Forecast Offices), but there are still some identifiable errors in the published data. In particular, from one forecast horizon to the next, there are a small number of observations that have a change in forecast value of

Figure A1: Comparison of Example Raw Gridded Forecast Data and County-level Data



(a) Gridded Raw Data (b) County Area Average (c) Population-weighted

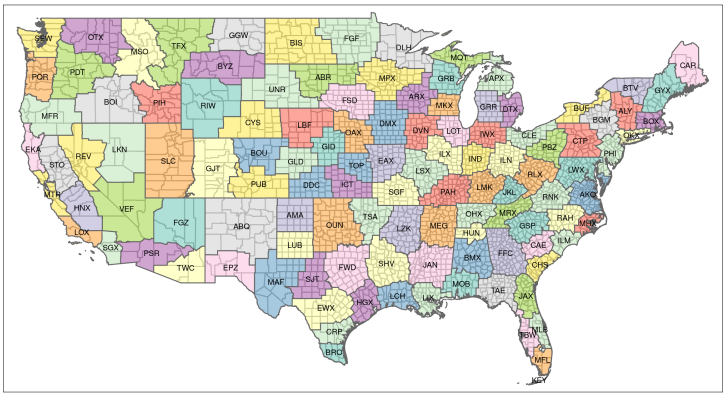
Notes: The maps show the raw, gridded forecast data in panel (a) and the corresponding county-level area and population-weighted average forecasts in panels (b) and (c) respectively. The maps are an example from one day and forecast horizon: the 1-day-ahead forecast for September 9, 2006.

exactly -17.4999°C. When these errors occur, they only appear at one forecast horizon, so we use adjacent forecast horizons to interpolate the erroneous value. In the primary results, we Winsorize the forecast errors, so this data cleaning step has minimal effect on the estimates.

Fifth, we match the timing conventions in the forecast and weather data. The NWS typically uses a noon to noon UTC convention for daily temperature forecasts. Minimum temperatures are forecasted for the nighttime (midnight UTC day to noon UTC day t or 7 p.m day $t-1$ to 7 a.m. day t EST). Maximum temperatures are forecasted for the daytime (noon UTC to midnight UTC). PRISM also typically follows this timing convention, but not as strictly. To match the timing conventions between the two datasets, for maximum and minimum temperatures separately we regress realized temperature on the day t 1-day-ahead forecast and the day $t-1$ 1-day-ahead forecast. For maximum temperature, we find that the day t forecast is sufficient (the day $t-1$ forecast does not predict the realization conditional on the day t forecast). For minimum temperature, we find that both days' forecasts are predictive, with the day t forecast being about twice as predictive as the day $t-1$ forecast. We therefore construct a time-corrected day t minimum temperature forecast that is the weighted average of the original day t and day $t-1$ forecasts with weight $2/3$ on the day t forecast and $1/3$ on the day $t-1$ forecast. The time-corrected forecast does exhibit forecast sufficiency.

After creating the daily, county-level dataset, we merge counties with identifiers for their NOAA County Warning Area (CWA). CWAs are collections of counties, and the local NWS Weather Forecasting Office (WFO) is responsible for generating forecasts for the CWA. The map of counties and CWAs is shown in Figure A2.

Figure A2: Map of NOAA County Warning Areas (CWAs)



Notes: The map shows (in colored areas with black outlines) the geographic boundaries of County Warning Areas (CWAs), the collection of counties for which a given NWS Weather Forecasting Office is responsible for creating forecasts. State borders are shown in gray, thinner lines. CWAs are typically composed of one or more counties and can cross state borders. There are 116 CWAs in the continental U.S. Some counties are part of multiple CWAs, and in those cases, we assign the county a CWA ID composed of each CWA that it is in. The end result is a many to 1 mapping of all continental U.S. counties to 130 CWAs or CWA groups.

A.2 Air Conditioning

We generate new predictions of air conditioning take-up at the county-by-year level. We begin with individual-level restricted access, biennial American Housing Survey (AHS) data from a Census Research Data Center that contains household-level information on air conditioning (AC) availability and demographic and household information. We link these data with county-by-year climatic characteristics from 1999 to 2020, from Schlenker and Roberts (2009). The AHS sample provides details at the household level that our model uses to predict AC penetration, while partially pooling using other households in the state to improve predictive fit.

We use a multi-step process to select the best model to predict AC availability at the household level. First, we consider all variables shared by the AHS and the public version of the ACS plus weather variables (annual rainfall; annual average temperature; annual maximum temperature; annual minimum temperature; average, max, and min summer temperature; and the annual standard deviation of daily temperature) as possible predictors. The full sample of data is then split, with 1/3 acting as the model testing sample and 2/3 as the hold-out, final prediction sample which helps avoid overfitting or the need for strong sparsity assumptions (Chernozhukov et al., 2018). The samples are blocked to ensure that all states and years are represented in both samples.

Candidate models are evaluated through 5-fold cross validation on the model testing

sample, again blocked at the state-year level. Within each cross-validation step, the set of chosen predictors is constructed by taking combinations of the possible predictors and fitting linear multilevel models in R with the `lme4` package (Bates et al., 2015). We model individuals as members of states, so that observations in relatively less sampled states are more strongly pooled toward their group mean to avoid fragile or high-variance out-of-sample predictions.

The first set of candidate models starts by fitting univariate models. We sequentially add each additional potential predictor, keeping the resulting model that achieves the best performance in terms of the Akaike information criterion (AIC). The addition of variables stops if the AIC increases or if a model fails to converge, in which case the last converging model is selected as a candidate. In a second set of models, the set of predictors is chosen through lasso (Tibshirani, 1996) and then a model is fit using `lme4`. These candidate models are then fit with time- and location-varying intercepts and slopes for each combination of up to three of the predictors. The final model is selected from the pool of best-performing models based on out-of-sample predictive fit across the cross-validation folds. The hold-out-sample is then used to estimate the final coefficients for disclosure. We fit the final model on 523,000 observations. We obtain an in-sample root mean squared error (RMSE) of 0.07696 with an average AC penetration rate at the county level of 0.8719.

We then bring these estimated coefficients to individual-level Integrated Public Use Microdata Series (IPUMS USA) data 1% sample household data from the American Community Survey (ACS) from 2005 to 2017. These ACS public use microdata files are geographically identified at the Census Public Use Microdata Area (PUMA). We standardize them by converting them to 2010 PUMA definitions using a 2000 to 2010 PUMA crosswalk by IPUMS. We similarly convert the NOAA climatological data at the county by year level to the PUMA by year level using the Census county to 2010 PUMA crosswalk. We predict household-level AC take up using estimated coefficients from the AHS data applied to our ACS data sample. Finally, we predict AC take-up at the county-by-year level by converting PUMA estimates back to counties: for counties that have a direct match with a PUMA, these estimates are directly applied, and for counties that match to multiple PUMAs, the weighted average of AC take up is estimated for the county-by-year level.

One version of our final data are these direct county-by-year AC take-up estimates, however we note that these unadjusted estimates do not monotonically increase over time as one may theoretically expect. Thus, in a second final estimate, we smooth the county-by-year estimates for each county based on a linear regression of the predicted county-by-year estimates for a given county. We force the overall trend to be (weakly) monotonically increasing. We estimate an average take-up of about 89.2%, broadly consistent with the average take-up of 87.2% in the original AHS data.

A.3 American Time Use Survey

American Time Use Survey data come from the Bureau of Labor Statistics (BLS) and are available at https://www.bls.gov/tus/. The data structure is a repeated cross section. Individuals who have taken part in the CPS are invited to complete a time use diary for a single day's time use. The sample is gathered uniformly throughout the year and across the country. About 10,000 to 12,000 individuals participate each year.

Data are geocoded at a variety of different levels depending on the population density in the location. All observations contain state-level geocoding. For high-density locations, geocoding is at the county level. For intermediate densities, one can use CPS records to geocode the observations at the CBSA level. Details on the geocoding process can be found in Gibson and Shrader (2018). We match individuals to weather and forecast records aggregated to their finest level of geocoding. For clustering, we assign each individual to a WFO either using their county or the WFO that covers the most area in their state.

A.4 Electricity Demand

Electricity demand data come from the US Energy Information Agency (EIA) form EIA-861M and are available at https://www.eia.gov/electricity/data.php. The dataset "Retail sales of electricity to ultimate customers - Monthly" contains monthly, state-level electricity consumption (MWh) and prices (cents/kWh) for residential, commercial, industrial, and other users. We combine the electricity data with weather and forecast data by aggregating the latter to the state-month level. Starting with the county-level data used in the main analysis, we calculate the number of days per month that a state experiences and is forecasted to experience weather in temperature bins (5°C for realized temperature and cold, cool, warm, and hot bins for forecasted temperature, to match Table 1). Rainfall is summed over days in the month and units are converted to meters per month for legibility. Forecast errors are averages of county-level errors. All of these calculations are weighted by county-level population. Finally, we merge the dataset with state-level population from the NIH Surveillance, Epidemiology, and End Results program.

B Additional Theoretical Analysis

B.1 Second-order condition

The following lemma gives sufficient conditions for the second-order condition to hold:

Lemma 2 (Second-Order Condition) *If Assumption 1 holds, $E_{T|f}[h(T, A^*(f))] \leq \epsilon$, and $h_{AA}(T, A^*(T)) \geq 0$, then the second-order condition holds around $A^*(f)$ as ϵ goes to 0.*

Proof. Differentiating the right-hand side of (2) with respect to A, applying Assumption 1 and Lemma 1, and letting $E_{T|f}[h(T, A)]$ be small, the second-order condition holds around $A^*(f)$ if $-C''u' + [C']^2u'' - E_{T|f}[h_{AA}](u - v) < 0$, which holds if $E_{T|f}[h_{AA}] \geq 0$. ■

Given that an individual's daily mortality risk is not generally large, the second-order condition should hold in our empirical application as long as the hazard function is not too concave in actions.

B.2 Identifying properties of the hazard function

Here we show how estimating which type of adaptation environment holds is informative about properties of the hazard function.

Begin by establishing that actions are sensitive to forecasts.

Assumption 3 (Constant VSL) *Around forecast f, the VSL is constant.*

Lemma 3 (Actions Respond to Forecasts) *If Assumptions 1 through 3 hold and the second-order condition for optimality of actions holds when the forecast is f, then* $\lim_{\epsilon \to 0} A^{*\prime}(f) \propto -h_{AT}(f, A^*(f))\, VSL$.

Proof. Applying the implicit function theorem to (4) and using the second-order condition, $A^{*\prime}(f) \propto -\frac{\partial E_{T|f}[h_A((T, A^*(f))]}{\partial f} VSL(f) - E_{T|f}[h_A((T, A^*(f))]\, VSL'(f)$. Using (5) and $VSL'(f) = 0$ (and thus dropping the argument of VSL), this becomes

$$A^{*\prime}(f) \propto -\left\{ h_{AT}(f, A^*(f)) + \frac{1}{2} h_{ATTT}(f, A^*(f)) Var[T|f] + \frac{1}{2} h_{ATT}(f, A^*(f)) \frac{\mathrm{d}Var[T|f]}{\mathrm{d}f} \right\} VSL.$$

The lemma follows from applying Assumptions 1 and 2. ■

Actions are sensitive to forecasts when the marginal effect of temperature on mortality risk depends on the actions chosen (i.e., when $h_{AT} \neq 0$).

Now consider what we learn about the hazard function:

Lemma 4 *If Assumptions 1 through 3 hold at $f = T$ with ϵ arbitrarily small and the second-order condition for optimality of actions holds when the forecast is $f = T$, then adaptation is appropriate if and only if* $\lim_{\epsilon \to 0} h_{AA}(T, A^*(T)) > 0$.

Proof. Using Lemmas 1 and 3, equation (6) implies that $\lim_{\epsilon \to 0} \mathrm{d}^2 h(T, A^*(T+e))/\mathrm{d}e^2|_{e=0} > 0$ if and only if $\lim_{\epsilon \to 0} h_{AA}(T, A^*(T)) > 0$. And from Definition 1, adaptation is appropriate if and only if $\mathrm{d}^2 h(T, A^*(T+e))/\mathrm{d}e^2|_{e=0} > 0$. ■

If we estimate that mortality risk is convex in forecast errors, then we can use Lemma 4 to conclude that the hazard is convex in actions around accurate forecasts. But if we instead estimate that mortality risk is not convex in forecast errors, then we can conclude that the hazard is linear or concave in actions around accurate forecasts.

B.3 Proof of Proposition 2

Second-order approximate $V(f)$ around $f = T$ inside $\bar{V}(T)$:

$$\bar{V}(T) \approx V(T) + \frac{1}{2} V''(T)\, Var[e|T]. \tag{A-1}$$

Therefore:

$$\frac{\mathrm{d}\bar{V}(T)}{\mathrm{d}Var[e|T]} \approx \frac{1}{2} V''(T), \tag{A-2}$$

with the approximation becoming exact when Assumption 1 holds with ϵ small. Second-order approximating $h(f - e, A^*(f))$ around $e = 0$ yields:

$$E_{e|f}[h(f - e, A^*(f))] \approx h(f, A^*(f)) + \frac{1}{2} h_{TT}(f, A^*(f))\, Var[e|f],$$

with the approximation again becoming exact when Assumption 1 holds with ϵ small. Substitute into $V(T)$ and apply Assumption 2 in order to hold $Var[e|T]$ constant in a neighborhood of T:

$$\begin{aligned} V'(T) = &- \left[h_T(T, A^*(T)) + \frac{1}{2} h_{TTT}(T, A^*(T)) Var[e|T] \right] \left[u(w - C(A^*(T))) - v(w - C(A^*(T))) \right] \\ &+ \frac{\mathrm{d}V(T)}{\mathrm{d}A} A^{*\prime}(T). \end{aligned}$$

Differentiating again yields:

$$\begin{aligned}
V''(T) = &- \left[h_{TT}(T, A^*(T)) + \frac{1}{2}h_{TTTT}(T, A^*(T))Var[e|T]\right] [u(w - C(A^*(T))) - v(w - C(A^*(T)))] \\
&- A^{*\prime}(T) \left[h_{AT}(T, A^*(T)) + \frac{1}{2}h_{ATTT}(T, A^*(T))Var[e|f]\right] \\
&\quad [u(w - C(A^*(T))) - v(w - C(A^*(T)))] \\
&+ A^{*\prime}(T)\, C'(A^*(T)) \left[h_T(T, A^*(T)) + \frac{1}{2}h_{TTT}(T, A^*(T))Var[e|T]\right] \\
&\quad [u'(w - C(A^*(T))) - v'(w - C(A^*(T)))] \\
&+ \left.\frac{\mathrm{d}\frac{\mathrm{d}V(f)}{\mathrm{d}A}A^{*\prime}(f)}{\mathrm{d}f}\right|_{f=T}.
\end{aligned}$$

The final line vanishes because the first-order condition must hold at all f. Substitute from (4) and then from (5),

$$\begin{aligned}
V''(T) \approx &- \left[h_{TT}(T, A^*(T)) + \frac{1}{2}h_{TTTT}(T, A^*(T))Var[e|T]\right] [u(w - C(A^*(T))) - v(w - C(A^*(T)))] \\
&- A^{*\prime}(T) \left[h_{AT}(T, A^*(T)) + \frac{1}{2}h_{ATTT}(T, A^*(T))Var[e|f]\right] \\
&\quad [u(w - C(A^*(T))) - v(w - C(A^*(T)))] \\
&- A^{*\prime}(T) \left[h_A(f, A^*(f)) + \frac{1}{2}h_{ATT}(f, A^*(f))Var[T|f]\right] VSL(T) \\
&\quad \left[h_T(T, A^*(T)) + \frac{1}{2}h_{TTT}(T, A^*(T))Var[e|T]\right] [u'(w - C(A^*(T))) - v'(w - C(A^*(T)))].
\end{aligned}$$

Using Assumption 1 and Lemma 1,

$$\lim_{\epsilon \to 0} V''(T) = - \lim_{\epsilon \to 0} \left[h_{TT}(T, A^*(T)) + A^{*\prime}(T)\, h_{AT}(T, A^*(T))\right] [u(w - C(A^*(T))) - v(w - C(A^*(T)))].$$

Observe that

$$\frac{\mathrm{d}}{\mathrm{d}e}\frac{\partial h(T, A^*(T + e))}{\partial T} = h_{AT}(T, A^*(T + e))\, A^{*\prime}(T + e),$$

which, from Lemma 3, is strictly negative if Assumption 3 and the second-order condition for optimality of actions both hold when the forecast is f. We then have:

$$\lim_{\epsilon \to 0} V''(T) = -\lim_{\epsilon \to 0} \left[\frac{\partial^2 h(T, A^*(T+e))}{\partial T^2}\bigg|_{e=0} + \frac{\mathrm{d}}{\mathrm{d}e} \frac{\partial h(T, A^*(T+e))}{\partial T}\bigg|_{e=0} \right] \left[u(w - C(A^*(T))) - v(w - C(A^*(T))) \right].$$

Substituting into (A-2), we find:

$$\lim_{\epsilon \to 0} \frac{\mathrm{d}\bar{V}(T)}{\mathrm{d}Var[e|T]} \approx -\lim_{\epsilon \to 0} \frac{1}{2} \left[\frac{\partial^2 h(T, A^*(T+e))}{\partial T^2}\bigg|_{e=0} + \frac{\mathrm{d}}{\mathrm{d}e} \frac{\partial h(T, A^*(T+e))}{\partial T}\bigg|_{e=0} \right] \left[u(w - C(A^*(T))) - v(w - C(A^*(T))) \right]. \tag{A-3}$$

Using (A-1), observe that, under Assumption 1,

$$\lim_{\epsilon \to 0} \frac{\mathrm{d}\bar{V}}{\mathrm{d}w} = E_{T|f}[1 - h(T, A)]\, u'(w - C(A^*(T))) + E_{T|f}[h(T, A)]\, v'(w - C(A^*(T))). \tag{A-4}$$

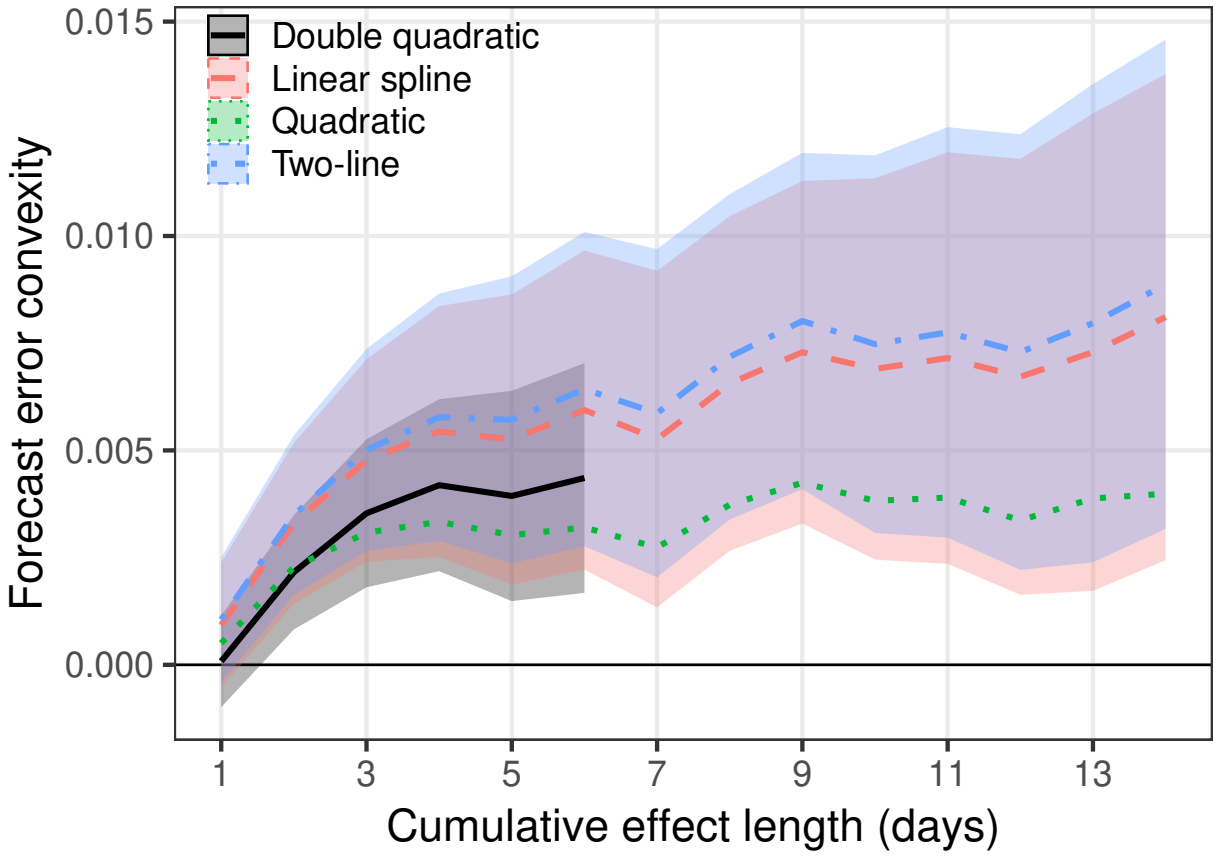
Plugging (A-3) and (A-4) into the definition of $WTP(T)$ in (8) and substituting from (3) yields the expression in the proposition.

C Testing Assumptions Underlying Forecast Benefit Estimates and Robustness to Functional Form

Section 6 shows the benefit of forecast improvements based on the 6-day cumulative effect of forecasts on mortality. One assumption underlying those results is that the effect of forecasts on mortality is persistent. Previous research on the effect of realized temperature on mortality has shown that there can be dynamic effects over the days following a temperature realization (Deschênes and Moretti, 2009, Heutel et al., 2021). Figure A3 shows the average marginal effect across the full temperature distribution of a more erroneous forecast (the same value reported in Column 5, row 3 of Table 3) using a two-week distributed lag model.[41] The results show that on the day the forecast arrives, the marginal effect is roughly 0.001 deaths per 100,000 people. This rises to around 0.002 after just one additional day, continues to rise through day 3, then stays roughly stable after that point. The value is significantly positive over all but the first day. This stability of the estimate supports our assumption of a persistent effect on mortality.

[41] Examining longer ranges of cumulative effects is computationally infeasible using a standard distributed lag model given the high dimensionality of the estimating equation.

Figure A3: Testing Counterfactual Persistence Assumption: Cumulative Effects



Notes: Shows estimates of the cumulative average convexity of mortality with respect to forecast error based on regressions estimated using Equation (10) ("Two-line"), a variant of that equation using a linear spline ("Linear spline"), a variant using a quadratic ("Quadratic"), and the baseline value estimating equation (14) ("Double quadratic"). All are fit to the baseline data and use 14 lags except the double quadratic. The shaded area shows the 95% confidence interval based on standard errors clustered at the CWA level.

A second assumption underpinning Table 3 is that the estimated forecast-mortality relationship holds under the counterfactual forecast. We generate descriptive evidence on this assumption by looking at how the effect varies by average forecast quality (measured by RMSE) in the sample. Results are shown in Figure 4. These estimates are non-causal, as locations with more accurate forecasts may differ from locations with less accurate forecasts in unobserved ways.[42] In each forecasted temperature bin, forecasts are more valuable in locations that have more accurate forecasts on average. If anything, this result suggests that our estimates are lower bounds on the mortality benefit of improved forecasts because more accurate forecasts are associated with stronger responses to forecasts.

For robustness, we assess the different estimates of lives saved that we find when estimating with different functional forms. In particular, we use the estimates of equation (10) reported in Table 1, a version that uses a linear spline rather than the two-line formulation, and a quadratic over forecast errors but not over temperature.[43]

The cumulative convexity derived from each of these equations is shown in Figure A3. The figure shows the average convexity across all temperature bins. For each functional form, the convexity rises for about 4 days then stabilizes. For the more parsimonious functional forms that permit longer cumulative effects to be computationally tractable, this stability is exhibited through two weeks with no indication of changing. Overall, the two linear models (two-line and linear spline) exhibit the largest average convexity. The two quadratic models exhibit slightly smaller average convexity, though still within the confidence bands of the linear models.

D Sensitivity to Breakpoint Choice for Two-line Test

The Simonsohn (2018) two-line procedure involves first selecting a breakpoint then fitting an interrupted regression on either side of that breakpoint. For selecting the breakpoint, Simonsohn proposes the "Robin Hood algorithm." The algorithm "donates" points from the more precisely estimated side of the interrupted regression to the weaker side so that overall power of the test is improved. In the original algorithm from Simonsohn (2018), the researcher first fits a flexible function relating the left-hand and right-hand side variables. This fit can be done using a nonparametric or semiparametric procedure. The researcher then

[42] There could be omitted variables that could cause forecasts to be more accurate and also cause errors to matter more (e.g., experiencing hot weather more frequently), and there could be selection in which places get more accurate forecasts (e.g., radars or skilled meteorologists may be directed to places where the National Weather Service believes that forecasts are more valuable). We adjust for observable confounders by including them in the regression, but unobservables might still confound the relationship.

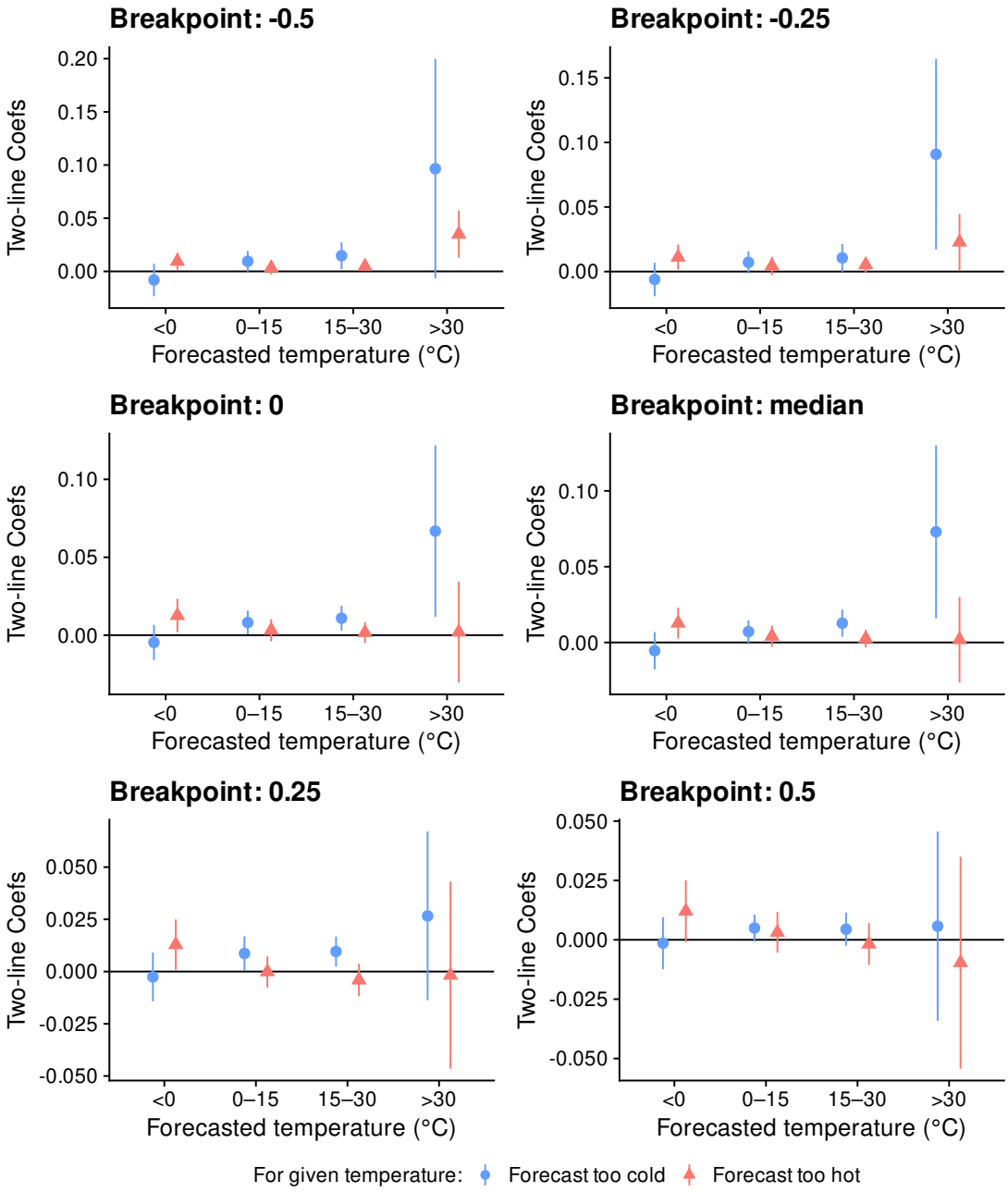
[43] We also estimated net benefits using a flexible polynomial specification rather than bins of forecasted temperature. The overall WTP estimates are highly similar in both cases, and the results are available upon request.

generates fitted values and chooses an initial candidate breakpoint, x_0, based on the most extreme fitted value. A range of additional candidate breakpoints is chosen by considering all x values within 1 standard error of the most extreme fitted value. This set of candidate breakpoints has upper bound x_H and lower bound x_L.

Next, the researcher fits an interrupted regression using the initial, candidate breakpoint x_0. The interrupted regression fit will lead to two initial slope estimates with associated standard errors $\hat{s}_H$ and $\hat{s}_L$ for the estimate above and below the breakpoint respectively. The final breakpoint is chosen by shifting the breakpoint to increase the number of points given to the less precisely estimated slope. Simonsohn proposes that the final breakpoint be $(\hat{s}_{imp}/(\hat{s}_L + \hat{s}_H) * 100$ percent of the way toward the edge of the boundary of candidate breakpoints, where $\hat{s}_{imp}$ is the standard error of the relatively less precise estimate.

In our setting, we have strong *a priori* reasons for preferring a breakpoint around 0 or median error, based on the theoretical analysis in Section 2. We thus use a median breakpoint for our main results. Figure A4 shows the estimated marginal effects using a range of different breakpoints. The breakpoint is indicated in the title of each panel. The results are very similar to the baseline (median breakpoints) when using a breakpoint of 0 because the median forecast error is close to 0 for all forecasted temperature bins. Results are close to the baseline results for the moderately negative and positive breakpoints (-0.25 and 0.25) although the "forecast too cold" coefficient in the hot temperature bin is not significant in the latter case. Results are weakest when the breakpoint is 0.5 (bottom right panel), a point that is half of a standard deviation away from the median error in most bins.

Figure A4: Robustness and Sensitivity Checks



Notes: The figures show sensitivity to the choice of breakpoint in estimating Equation 10. All models use the same functional form, lag length, and controls as the baseline results. The lines are 95% confidence intervals based on standard errors clustered at the CWA level. The panel titles indicate the forecast error breakpoint used in the panel. The baseline results are those that use median breakpoints and correspond to the results in Table 1.

E Additional Robustness Checks

Figure A5 shows robustness and sensitivity checks for the main results shown in Table 1. The red triangles show the marginal effect of a forecast that is too hot and the blue circles show the marginal effects for a forecast that is too cold. The lines are 95% confidence intervals. The figures show results for forecasted cold temperatures (panel a), cool temperatures (panel b), warm temperatures (panel c), and hot temperatures (panel d). Each panel is broken into 3 sections. The first section varies the standard error clustering; the second section varies the non-weather related controls; and the third section varies the realized weather, pollution, and forecast controls.

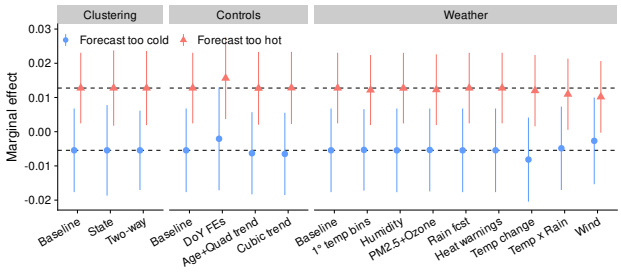
In the clustering section, the baseline estimate is clustered at the CWA level. "State" clusters at the state, and "Two-way" clusters at the county and year level. The results are typically similar across all of these clustering schemes.

In the controls section, baseline estimate includes controls for 5°C bins of realized temperature, lags of indicators for above median precipitation, date fixed effects, and county-by-month fixed effects interacted with linear time trends. "DoY FEs" replaces the month fixed effects with day-of-year fixed effects. "Age+Quad trend" add a quadratic year trend interacted with county-by-month fixed effects and month fixed effects interacted with four population age indicators. This exactly matches the control set used in Barreca et al. (2016) but adapted to our county-level dataset rather than a state-level dataset. "Cubic trend" adds cubic trends interacted with county-by-month fixed effects. The day of year fixed effects have the strongest effect on the estimates among this set of checks.

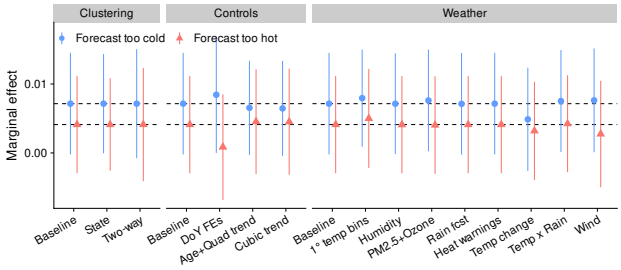
The third section varies controls for weather, pollution, and other forecasts. Unless otherwise stated, the controls for weather are included using 8 bins equally spaced using percentiles of the data. "1° temp bins" replaces the 5°C bins with finer controls for realized temperature. "Humidity" controls for the relative humidity. "PM2.5+Ozone" controls for county-level ambient ozone and $PM_{2.5}$ concentrations from EPA's RSIG Fused Air Quality Surface Using Downscaling (FAQSD) files available here: `https://www.epa.gov/hesc/rsig-related-downloadable-data-files`. "Rain fcst" includes 1-day-ahead rainfall forecasts binned to match the realized rainfall controls, "Temp change" includes the change in temperature between day t and $t-1$, "Temp x Rain" interacts the realized rain and temperature controls, and "Wind" controls for both direction and speed of wind as measured by NOAA's North American Regional Reanalysis (NARR) available here: `https://psl.noaa.gov/data/gridded/data.narr.html`.

Like PRISM, the NARR dataset combines individual weather observations with a model (in this case, the NCEP Eta weather model) to produce weather measures on a consistent grid across the U.S (Mesinger et al., 2006). The grid has a spatial dimension of roughly 32km,

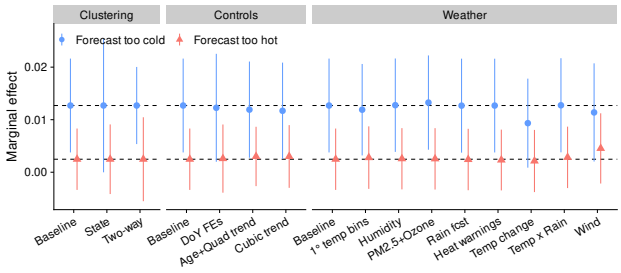
Figure A5: Robustness and Sensitivity Checks



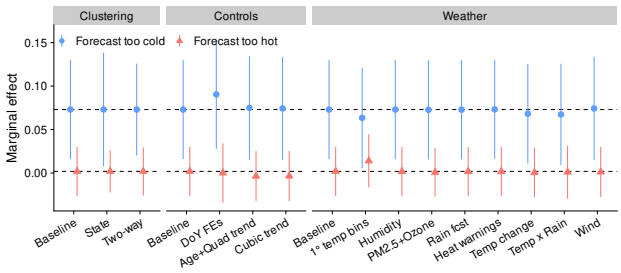
(a) $< 0°$



(b) 0–15°



(c) 15–30°



(d) $> 30°$

Notes: The figures show robustness and sensitivity checks on the main results reported in Section 4. All models use the same functional form and lag length as the baseline results. The lines are 95% confidence intervals based on standard errors clustered at the CWA level. The labels indicate the change or addition. For comparison, "Baseline" reproduces the baseline estimate, with controls described in Section 4.1. In all cases, added weather variables are controlled for non-parametrically using quantile bins, and 6 lags of the bins are included to match the lag length of the forecast error. Note that wind and rainfall forecasts are only available for a subset of the observations, so the sample changes.

and we take a spatial average of the values in each county to match our estimation sample.[44] Because the wind data grid is coarser than the PRISM grid, we lose some county observations when we include wind. The sample also changes for the rainfall point forecast because rainfall point forecasts were included in the NDFD at a later date than the temperature point forecasts used in the main analysis.

Among the weather controls, the largest effects come from using finer temperature bins, including temperature changes, and including wind. Finer temperature bins reduce the effect of negative forecast errors during hot periods and increase the effect of forecast errors that are too hot. Temperature changes reduce the effect of negative forecast errors during both cool and warm periods. Including wind moves marginal effects toward each other during both cold and warm periods.

F Counterfactual Approximation Quality

In Section 2, we derive marginal conditions for forecast value. These conditions are second-order approximations to the value of a change in forecast error distribution. For a marginal change in forecast error standard deviation, this follows from Proposition 1,

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{\mathrm{d}\bar{h}(T)}{\mathrm{d}\sigma_{e|T}}n(T) =& \lim_{\epsilon\to 0}\frac{\mathrm{d}\bar{h}(T)}{\mathrm{d}Var[e|T]}\frac{\mathrm{d}Var[e|T]}{\mathrm{d}\sigma_{e|T}}n(T)\\ =& \lim_{\epsilon\to 0}\frac{\mathrm{d}^2h(T,A^*(T+e))}{\mathrm{d}e^2}\bigg|_{e=0}\sigma_{e|T}n(T) \qquad \text{(A-5)}\end{aligned}$$

where $\sigma_{e|T}$ is the standard deviation of forecasts (and forecast errors) given temperature T. For a $X\times 100$ percent reduction in forecast error standard deviation, the change in mortality is scaled by $(1-(1-X)^2)\sigma^2_{e|T}$ instead of $2\sigma_{e|T}$.

For a discrete change in the forecast error distribution, the value is given by the difference in expected value under the counterfactual and actual distributions. The approximation will be accurate if the distribution of errors is close to normal or if the mortality hazard function is approximately quadratic in forecast errors. The approximation is practically useful because it is faster to compute. Table A1 compares the estimated counterfactual across all realized temperatures using both the approximation and a nonparametric estimate. One can see that the approximation accurately reproduces the results from the nonparametric estimator in this setting, with a difference in estimates of no more than 5%.

[44] Further details on the steps we follow to process the wind data can be found in Missirian (2020).

Table A1: Comparison of Counterfactual Approximation and Nonparametric Calculation for Monetized Lives Saved

Forecasted temperature:	(1) $< 0°$	(2) $0 - 15°$	(3) $15 - 30°$	(4) $> 30°$
Approximation method				
Nonparametric	110.923	768.558	1268.245	44.925
Approximation	116.476	789.543	1290.889	46.356
Approx./Nonpar.	1.05	1.027	1.018	1.032

Notes: The table compares estimates of the counterfactual lives saved from a 50% improvement in forecast error calculated using a nonparametric approximation (row 1) and second-order approximation (row 2). The counterfactual values correspond to the "50% improvement" row from Table 3.

G Heterogeneity by Demographics, Cause of Death, and Region

The CDC mortality records provide three dimensions of demographic information about the deceased individuals. They also list the cause of death. Figures A6a, A6b, A6c, and A7 show heterogeneity results along these different dimensions, based on estimates of equation (14) where the left-hand side variable has been replaced with mortality for the demographic or cause of death group listed in the figure. In all panels, the y-axis shows the annual lives saved per 100,000 people. In each of the figures, the top left panel shows the estimates for cold forecasted temperatures, the top right panels show it for cool forecasted temperatures, bottom left for warm, and bottom right for hot.

For sex (Figure A6a), the coefficients are almost always the same for both men and women. The one exception is forecasts that are too cold on days that are forecasted to be cold. This reduces mortality for women, if anything. This results in per capita lives saved that are similar for men and women at all but cold temperatures, where men have more positive benefit while women have a value near zero (although the difference is not statistically significant).

For age (Figure A6b), the strongest effects come from individuals older than 35. Point estimates are small for young people. In general, there is monotonically increasing per capita lives saved from forecast improvements as individuals get older for all temperature bins. The one exception is the cold bin, where individuals between ages 75 and 84 experience the highest per capita mortality reductions and all other groups show mortality reductions near zero. In unreported results, there is substantial heterogeneity in the effect within the 0 to 19 age group, with the largest point estimates for children between 1 and 5 years old, and a slightly negative point estimate for infants less than 1. In all cases, however, the confidence intervals

for these groups are extremely wide.

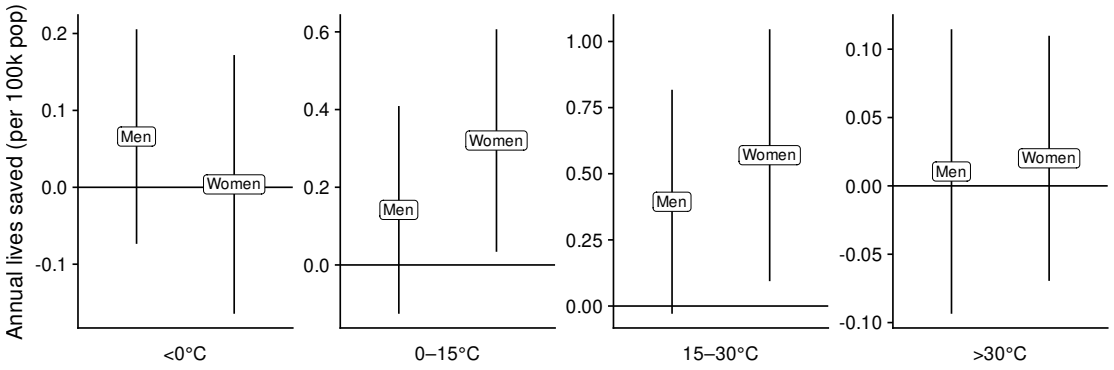
Figure A6c compares forecast effects by the race of the deceased, with separate estimates for individuals who identified as white, Black, or other races (Asian or Pacific Islander, and American Indian or Alaska Native). The effect of forecasts on mortality is substantially greater for white individuals than for all other individuals. Notably, in Figure A6c (the same estimates as reported in the body of the text in Figure 5), the point estimates indicate that forecasts across the temperature distribution have close to zero effect on mortality for all people of color.

Figure A7 shows estimates by cause of death. Causes of death reported on death certificates are subject to discretion by the individual filling out the death certificate, so all results should be taken as noisy and weakly informative. In terms of point estimates, the causes that are most significantly associated with forecasts are acute respiratory failure, accidents, cardiovascular disease, as well as other disease and "all other" which captures any cause of death that is not explicitly categorized.

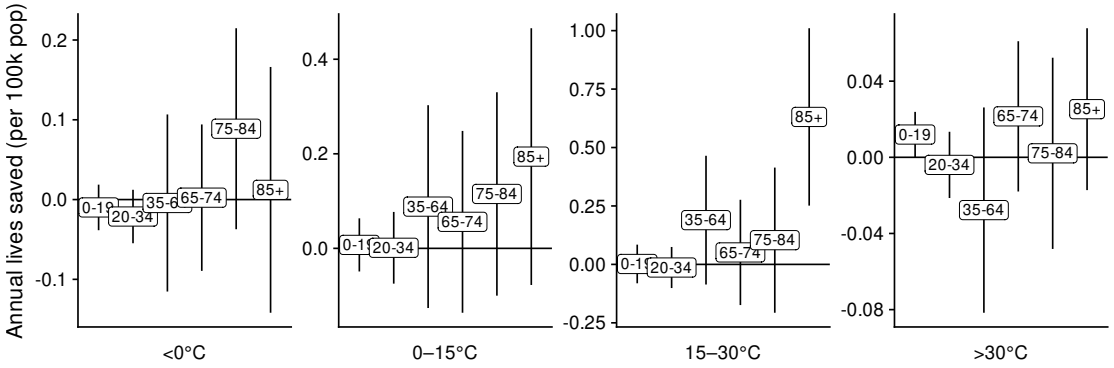
The strong associations with respiratory, cardiovascular, and cancer deaths is consistent with the findings on leading causes of death from temperature exposure (Deschênes and Moretti, 2009). The higher association with accidents is not found in studies of realized temperature and could be due to avoidance behavior engaged in by individuals to try to reduce their exposure to extreme weather.

The Figures A8 and A9 show the annual, lives saved per 100,000 people from marginal forecast improvements for the 9 NOAA climate regions (indicated on the x-axis of each figure). The points are estimates and the whiskers are 95% confidence intervals.

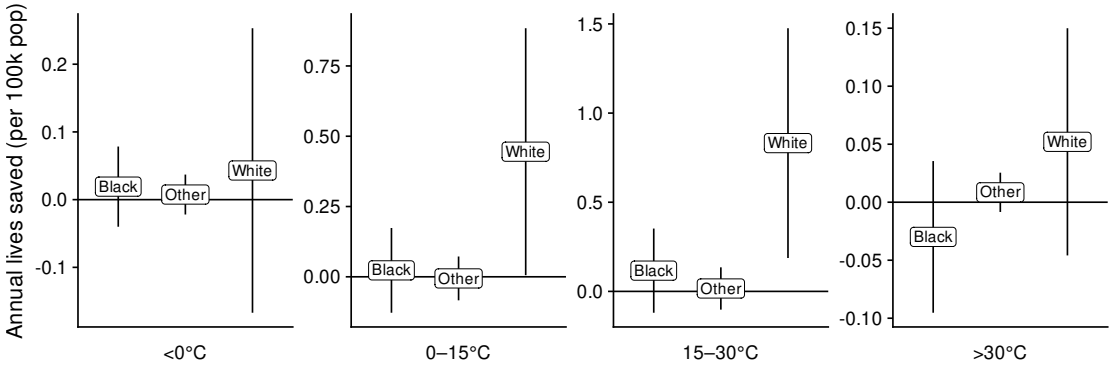
Figure A6: Heterogeneity by Demographics of the Deceased



(a) Sex



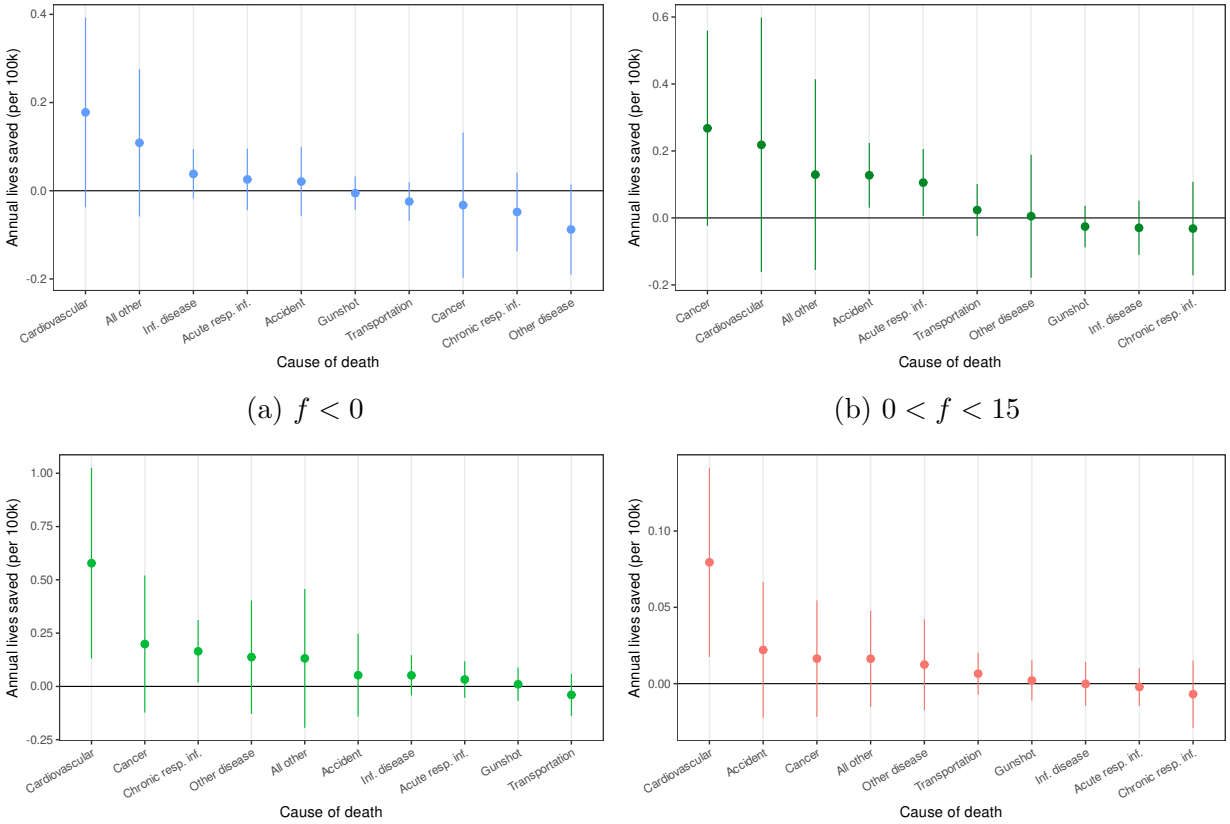
(b) Age



(c) Race

Notes: The figure shows the 6-day cumulative annual, lives saved per 100,000 people from a marginal decrease in 1-day-ahead forecast error. Estimates for each demographic category come from a separate model fit using Equation (14) on the baseline data where the dependent variable is mortality in the indicated demographic group. The range of forecasted temperature is indicated below each figure. Demographic categories are shown in the labelled points, which are also the point estimates derived from estimation. The lines are 95% confidence intervals based on standard errors clustered at the CWA level.

Figure A7: Heterogeneity by Cause of Death



(a) $f < 0$

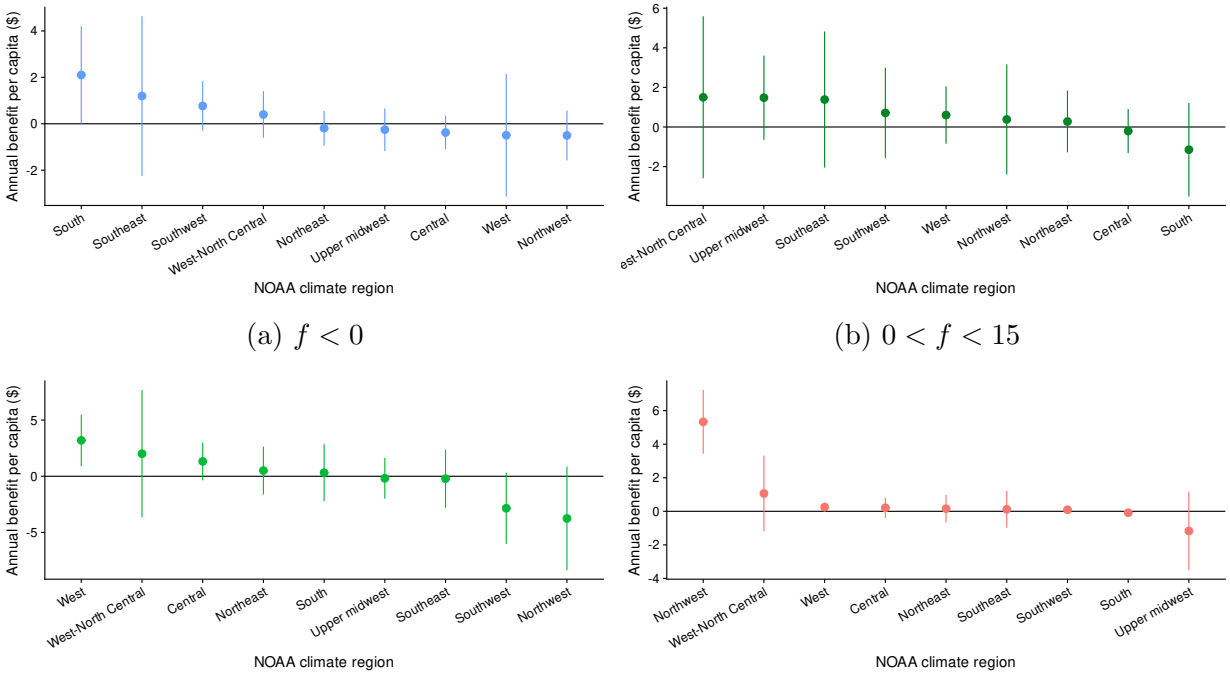
(b) $0 < f < 15$

(c) $15 < f < 30$

(d) $30 > f$

Notes: The figure shows the 6-day cumulative effect on annual lives saved per 100,000 people from a marginal reduction in forecast error. Estimates for each cause of death come from a separate model fit using Equation (10) on the baseline data. The cause of death is shown on the x-axis, and estimates are ordered by the effect size for negative forecast errors for $f < 0$°C and for positive forecast errors for all other panels. The range of forecasted temperature is indicated below each figure. Circles are the point estimates and the lines are 95% confidence intervals based on standard errors clustered at the CWA level.

Figure A8: Monetized Mortality Benefits of Marginal Forecast Improvement: Regional Heterogeneity by Forecasted Temperature



(a) $f < 0$

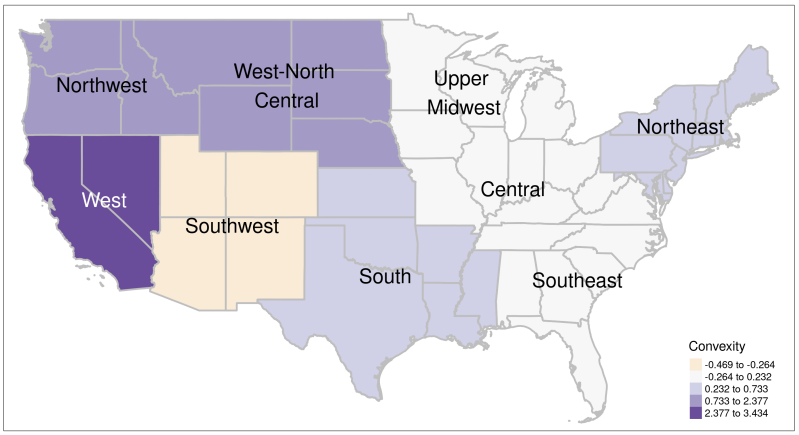
(b) $0 < f < 15$

(c) $15 < f < 30$

(d) $30 > f$

Notes: The figure shows the 6-day cumulative percent increase in the per-capita mortality rate, monetized using the VSL, from a 1°C increase in forecast mean absolute error based on a single model fit using Equation (10) on the baseline data. The forecast error variables are interacted with indicators for each NOAA climate region in the Continental U.S. The circles are the point estimates and the lines are 95% confidence intervals based on standard errors clustered at the CWA level. A map showing these effects spatially can be found in Figure A9.

Figure A9: Spatial Heterogeneity in Forecast Effect



Notes: The figure shows the 6-day cumulative effect on the mortality rate from a 1°C reduction in forecast absolute error based on Equation (10) estimated across NOAA climate regions. Darker purple colors indicate stronger benefits from forecast improvements, while lighter, orange colors indicate lower benefits.

H Longer-horizon Forecasts

The NWS issues point forecasts with horizons of up to 1 week. Table A2 shows 2-day cumulative effects when including both the 1-day-ahead forecast and longer-horizon forecasts in the estimation simultaneously. We focus on 2-day effects to simplify the interpretation when including multiple forecast horizons—looking only over 2 days means that the 1-day-ahead forecast is always the most recent, available information included in the regression. If there are adjustment costs that hamper individuals from acting on shorter-horizon forecasts, then longer-horizon forecasts can provide more adaptation benefits. This will show up as a convex relationship between mortality and the longer-horizon forecast, even conditional on the shorter-horizon forecast.

Table A2: Effects by Forecast Horizon: Convexity of Pooled Estimates

	(1) Mortality rate	(2) Mortality rate	(3) Mortality rate
1-day ahead convexity	0.0030***	0.0022***	0.0022**
	(0.0006)	(0.0006)	(0.0008)
3-day ahead convexity		0.0008**	
		(0.0004)	
6-day ahead convexity			0.0008***
			(0.0002)
Dependent var. mean	2.25	2.24	2.24
N	13,529,776	13,408,395	11,078,870
N Clusters	130	130	130

Notes The table shows 2-day cumulative effects from estimation of versions of Equation (10) that also include longer-horizon forecasts and use a quadratic specification rather than a two-line specification to capture nonlinearity in the effect of forecast errors on mortality. The dependent variable is the daily mortality rate per 100,000 people. The coefficients are the average marginal effect of a more erroneous forecast, pooled across all forecasted temperature bins. All models include the baseline model covariates and weighting. Standard errors, clustered at the CWA-level, are below each estimate. Significance: $p < .10$, ** $p < .05$, *** $p < .01$.

The results in Table A2 are consistent with longer-horizon forecasts providing additional value over-and-above the day-ahead forecasts. For both the 3- and 6-day forecasts, forecast errors have a significant, convex relationship with mortality. Notably, in Columns (2) and (3), the sum of the effects of the 1-day-ahead forecast and the 3- or 6-day ahead forecast are approximately equal to the effects of the 1-day-ahead forecast in Column (1), where the longer-horizon forecasts are not included. Forecasts at all horizons are highly correlated, so

the 1-day-ahead forecast effect in Column (1) (and in our other results) captures many of the benefits of all horizons of forecasts issued by the NWS.

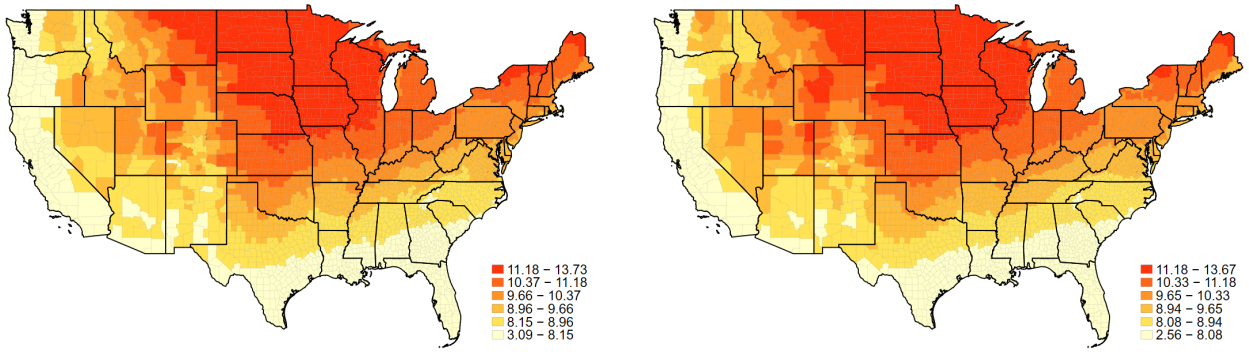
I Additional Figures and Tables

Table A3: Summary Statistics

Variable	Mean	S.D.	Observations
Daily all-cause mortality rate (per 100,000)	2.25	1.73	13,699,990
Average temperature (°C)	14.557	10.052	13,699,990
1-day-ahead avg. temperature forecast (°C)	14.515	9.964	13,699,990
1-day-ahead forecast error (°C)	-0.041	1.146	13,699,990
1-day error if $f < 0$ (°C)	-0.008	1.396	1,775,001
1-day error if $0 \leq f < 15$ (°C)	-0.007	1.229	5,261,602
1-day error if $15 \leq f < 30$ (°C)	-0.075	1.031	6,516,717
1-day error if $f \geq 30$ (°C)	0.060	0.905	146,670

Notes: The table shows summary statistics for the primary variables in the estimation sample, weighted by county population. The difference between average realized temperature and average forecasted temperature does not necessarily equal the average forecast error due to rounding.

Figure A10: Unconditional Variation in Temperature and Day-ahead Forecast

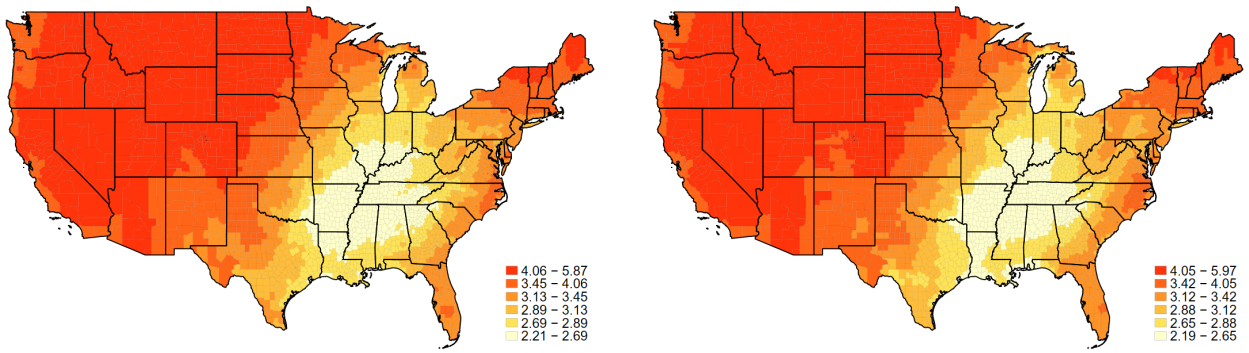


(a) St. dev. of average temperature

(b) St. dev. of 1-day-ahead forecast

Notes: The maps show the standard deviation of the unconditional average temperature (left panel) or the 1-day-ahead forecast of average temperature (right panel). For an indication of the identifying variation conditional on controls, compare these maps to the maps in Figure A11.

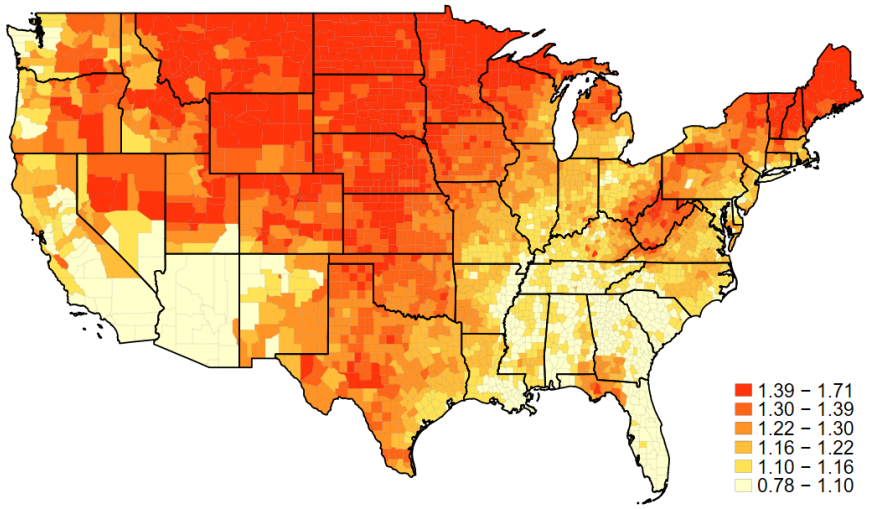
Figure A11: Residual Variation in Temperature and Day-ahead Forecast



(a) St. dev. of residual average temperature (b) St. dev. of residual 1-day-ahead forecast

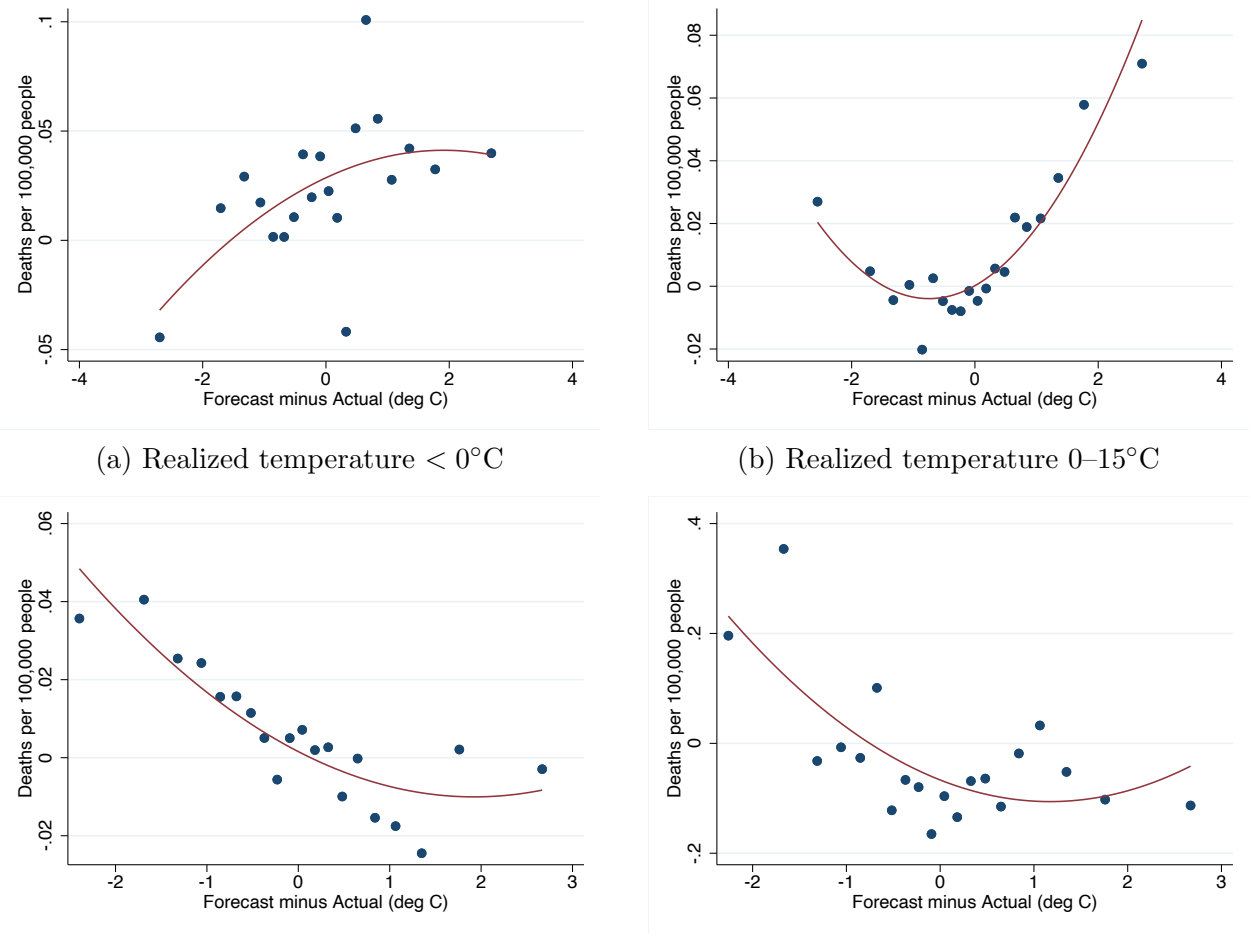
Notes: The maps show the standard deviation of the residuals from a regression of average temperature (left panel) or the 1-day-ahead forecast of average temperature (right panel) on all of the controls in the baseline regression specification (see Equation 10).

Figure A12: Spatial Variation in Forecast RMSE (Conditional on Baseline Controls)



Notes: The map shows the root mean squared error of the 1-day-ahead forecast for each county in the continental U.S. over the sample period. Redder values indicate higher average RMSE and yellower values indicate lower values. The values are all conditional on the baseline fixed effects and other control variables.

Figure A13: Raw Data Relationship Between 1-Day Ahead Forecast Error and Mortality for Days with Hot and Cold Realized Temperatures



(a) Realized temperature < 0°C

(b) Realized temperature 0–15°C

(c) Realized temperature 15–30°C

(d) Realized temperature > 30°C

Notes: The figures show bin-scatters of the relationship between forecast error from the 1-day ahead forecast and the daily mortality rate using the raw data residualized on realized temperature. The points are the average mortality rate within 20 quantiles of forecast error. Panel (a) shows the relationship when the expected temperature is cold (< 0°C), Panel (b) when cool (0–15°C), Panel (c) when warm (15–30°C), and Panel (d) when hot (> 30°C).

Table A4: Time Use: Quadratic Specification

	(1) Work	(2) Home prod.	(3) Leisure
$< 0°$ × Forecast error	4.01	-2.17	-1.84
	(4.20)	(4.18)	(6.95)
0 to $15°$ × Forecast error	0.35	-5.50***	5.16**
	(2.67)	(1.99)	(2.59)
15 to $30°$ × Forecast error	1.97	-1.01	-0.96
	(3.13)	(3.02)	(2.43)
$> 30°$ × Forecast error	-1.01	23.5*	-22.5*
	(10.2)	(13.9)	(11.8)
$< 0°$ × Forecast error^2	-5.30*	1.48	3.82
	(2.99)	(2.92)	(2.56)
0 to $15°$ × Forecast error^2	-1.06	1.13	-0.068
	(1.93)	(1.70)	(1.60)
15 to $30°$ × Forecast error^2	-5.46***	2.25	3.21*
	(1.91)	(1.71)	(1.88)
$> 30°$ × Forecast error^2	19.6**	6.42	-26.0***
	(8.99)	(7.07)	(9.59)
LHS mean	189.8	263.8	986.4
N	144,234	144,234	144,234
Clusters	100	100	100

Notes: The table shows estimates of the forecast error effect from a quadratic version of the time use regressions reported in Table 2. Additional covariates match the main results and are $5°$ bins for realized temperature, four bins for forecasted temperature, quartile bins for precipitation, as well as fixed effects for date, and location-by-month interacted with a linear time trend. Weighted by location population. Standard errors, clustered at the WFO level, are below each estimate. Significance: $p < .10$, ** $p < .05$, *** $p < .01$.

Table A5: Residential Electricity Demand: All Weather and Forecast Estimates

	(1)	(2)
	log electricity demand	
Temperature $< -10°$	0.013***	0.013***
	(0.0033)	(0.0034)
Temperature -10 to $-5°$	0.011***	0.011***
	(0.0026)	(0.0026)
Temperature -5 to $0°$	0.011***	0.011***
	(0.0026)	(0.0026)
Temperature 0 to $5°$	0.0077***	0.0077***
	(0.0015)	(0.0015)
Temperature 5 to $10°$	0.0045***	0.0046***
	(0.0016)	(0.0016)
Temperature 10 to $15°$	0.0032***	0.0032***
	(0.0012)	(0.0012)
Temperature 20 to $25°$	0.0051***	0.0051***
	(0.0013)	(0.0013)
Temperature 25 to $30°$	0.0087***	0.0087***
	(0.0021)	(0.0021)
Temperature $> 30°$	0.012***	0.012***
	(0.0025)	(0.0026)
Rain	-0.00029	-0.00021
	(0.00027)	(0.00027)
Rain × Rain	0.0000024	0.0000017
	(0.0000046)	(0.0000046)
Days $< 0°$	-0.00044	-0.00058
	(0.0013)	(0.0012)
Days 0 to $15°$	-0.00010	-0.00020
	(0.00039)	(0.00040)
Days $> 30°$	-0.0010**	-0.0012**
	(0.00046)	(0.00048)
Days $< 0°$ × Forecast error	0.00029	0.00023
	(0.00026)	(0.00025)
Days 0 to $15°$ × Forecast error	-0.00029	-0.00023
	(0.00020)	(0.00021)
Days 15 to $30°$ × Forecast error	0.00014	0.00016
	(0.00019)	(0.00019)
Days $> 30°$ × Forecast error	0.0014***	0.0013***
	(0.00033)	(0.00033)
Baseline controls	Yes	Yes
Log price	No	Yes
Observations	7104	7104
Clusters	48	48

Notes: The table shows all estimates from the regression reported in Column 1 of Table 2 plus an additional version of the regression that includes the log of the electricity price. Standard errors, clustered at the state level, are below each estimate. Significance: $p < .10$, ** $p < .05$, *** $p < .01$.

Table A6: Residential Electricity Demand: Quadratic Specification

	(1)	(2)
	log electricity demand	
Days $< 0°$ × Forecast error	0.000099	0.000030
	(0.00027)	(0.00026)
Days 0 to $15°$ × Forecast error	-0.00012	-0.000043
	(0.00024)	(0.00026)
Days 15 to $30°$ × Forecast error	0.0000047	0.000020
	(0.00018)	(0.00018)
Days $> 30°$ × Forecast error	0.00094**	0.00093**
	(0.00038)	(0.00038)
Days $< 0°$ × Forecast error^2	0.00038***	0.00039***
	(0.00010)	(0.00010)
Days 0 to $15°$ × Forecast error^2	-0.00033***	-0.00035***
	(0.000091)	(0.000088)
Days 15 to $30°$ × Forecast error^2	0.00033***	0.00034***
	(0.00012)	(0.00012)
Days $> 30°$ × Forecast error^2	0.0011*	0.0011**
	(0.00055)	(0.00051)
Baseline controls	Yes	Yes
Log price	No	Yes
Cost ($)	3,397,047	3,383,431
	(1,712,338)	(1,853,717)
N	7104	7104
Clusters	48	48

Notes: The table shows estimates of the forecast error effect from a quadratic version of the residential electricity demand regression reported in Table 2. Additional covariates are $5°$ bins for realized temperature, four bins for forecasted temperature, a quadratic for precipitation, and the log of the price for residential electricity, as well as fixed effects for year-month, and state-by-month interacted with a linear time trend. Weighted by state population. Standard errors, clustered at the state level, are below each estimate. Significance: $p < .10$, ** $p < .05$, *** $p < .01$.

Appendix References

Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2016). Adapting to climate change: The remarkable decline in the U.S. temperature-mortality relationship over the 20th century. *Journal of Political Economy 124*(1), 105–159.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using Lme4. *Journal of Statistical Software 67*(1), 1–48.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

CIESIN (2017). U.S. Census grids 2010. https://sedac.ciesin.columbia.edu/data/collection/gpw-v4. Accessed: 2018-08-30.

Deschênes, O. and E. Moretti (2009). Extreme weather events, mortality, and migration. *The Review of Economics and Statistics 91*(4), 659–681.

Gibson, M. and J. Shrader (2018). Time use and labor productivity: The returns to sleep. *Review of Economics and Statistics 100*(5), 783–798.

Heutel, G., N. H. Miller, and D. Molitor (2021). Adaptation and the mortality effects of temperature across us climate regions. *Review of Economics and Statistics 103*(4), 740–753.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, et al. (2006). North american regional reanalysis. *Bulletin of the American Meteorological Society 87*(3), 343–360.

Missirian, A. (2020). Yes, in your backyard: Forced technological adoption and spatial externalities.

Schlenker, W. and M. J. Roberts (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.

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science 1*(4), 538–555.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.