

# Forecast Accuracy Matters for Hurricane Damages

Andrew B. Martinez  
Office of Macroeconomic Analysis, US Department of the Treasury;  
GWU Research Program on Forecasting;  
Climate Econometrics at Nuffield College  
[andrew.martinez@treasury.gov](mailto:andrew.martinez@treasury.gov)

RPF Working Paper No. 2020-003

May 5, 2020

RESEARCH PROGRAM ON FORECASTING  
Center of Economic Research  
Department of Economics  
The George Washington University  
Washington, DC 20052  
<https://www2.gwu.edu/~forcpgm>

# Forecast Accuracy Matters for Hurricane Damages

Andrew B. Martinez\*

May 5, 2020

## Abstract

I analyze damages from hurricane strikes on the United States since 1955. Using machine learning methods to select the most important drivers for damages, I show that large errors in a hurricane's predicted landfall location result in higher damages. This relationship holds across a wide range of model specifications and when controlling for ex-ante uncertainty and potential endogeneity. Using a counterfactual exercise I find that the cumulative reduction in damages from forecast improvements since 1970 is about \$82 billion, which exceeds the U.S. government's spending on the forecasts and private willingness to pay for them.

Keywords: Adaptation, Model Selection, Natural Disasters, Uncertainty

*JEL classifications:* C51, C52, Q51, Q54

---

\*Office of Macroeconomic Analysis, US Department of the Treasury; GWU Research Program on Forecasting; Climate Econometrics at Nuffield College. Contact: andrew.martinez@treasury.gov. The views expressed here are my own and not those of the Treasury Department or of the U.S. Government. This research was supported in part by a grant from the Robertson Foundation (grant 9907422). Thanks to Jennifer L. Castle, Jurgen A. Doornik, Neil R. Ericsson, David F. Hendry, Luke P. Jackson, Felix Pretis, participants at the Conference on Econometric Models of Climate Change, the 6th International Summit on Hurricanes and Climate Change, Oxford Econometrics Lunch Seminar, and the VU Amsterdam Institute for Environmental Studies and to two anonymous reviewers for helpful comments and suggestions. Special thanks to Luke P. Jackson for help with processing the soil moisture data. Numerical results were obtained using OxMetrics 7.2 (OSX/U), PcGive 14.2 and R 3.4.3. All errors are my own.

# 1 Introduction

Damages from natural disasters in the United States, driven in large part by Hurricane Harvey and Hurricane Irma, reached a record high of \$313 billion in 2017 (NOAA NCEI, 2018). Hurricanes now account for seven of the top ten costliest disasters in the United States since 1980. They can have substantial impacts on local economic growth (Strobl, 2011), fiscal outlays (Deryugina, 2017), lending and borrowing patterns (Gallagher and Hartley, 2017), and on where people live and work (Deryugina et al., 2018). Given the severity of these impacts, it is important to understand what drives the destructiveness of these events.

While there are many determinants of damages, hurricanes are local events whose impacts are affected by individuals and businesses adaptation choices. Decisions about how and when to protect property or evacuate are made in advance and rely on forecasts of the storm; see Shrader (2018), Kruttli et al. (2019) and Beatty et al. (2019). However, hurricane forecasts, despite dramatic improvements, are far from perfect and can exhibit large and unexpected errors. These errors, even up to just a few hours ahead, can lead individuals in a disaster area to protect their property less than they would have otherwise and lead to higher damages.

In this paper I evaluate whether hurricane forecast accuracy matters for aggregate hurricane damages. I start by formulating an empirical model of damages with many determinants and then use model selection methods to determine the best damage model specification. Next, I exploit the natural variation in the forecast errors to quantify the impact of forecast accuracy on hurricane damages. Short-term forecast errors together with a handful of other variables, explain most of the variation in aggregate hurricane damages over the past sixty years. I find that a one standard deviation increase in the hurricane strike location forecast error is associated with up to \$9,000 in additional damages per household affected by a hurricane. This is about 18 percent of the typical post-hurricane flood insurance claim.

Little attention has been paid to the role that forecasts of natural disasters can play in reducing their destructiveness. For example Deryugina et al. (2018, p. 202) ignores the role of forecasts by claiming that Hurricane Katrina “struck with essentially no warning”. Others focus only on potential long-term impacts by arguing that improved forecasts can increase damages in the longer-term as people perceive declining risks and relocate to more

hurricane-prone areas; Sadowski and Sutter (2005) and Letson et al. (2007). I take a different approach by using a semi-structural partial-equilibrium framework to show that the cumulative damages prevented due to improvements in forecast accuracy since 1970 is about \$82 billion. Furthermore, the cumulative net benefit is between \$30 – 71 billion after subtracting estimates of what the U.S. federal government has spent on hurricane operations and research. This illustrates that improvements in the forecasts produce benefits beyond the well-documented reduction in fatalities and have outweighed the associated costs.

I also assess the importance of forecast uncertainty by adapting Rossi and Sekhposyan (2015)’s measure of uncertainty to allow for time-varying forecast-error densities. I decompose this measure into the ex-post forecast errors and its ex-ante standard deviation. This allows me to test whether errors in ex-ante beliefs about the storm or the strength of those beliefs play greater roles in altering damages from natural disasters. I find that it is the errors in ex-ante beliefs that matter most for hurricane damages.

The rest of the paper is structured as follows: Section 2 describes the modeling framework and section 3 describes the forecasts. The model selection methods are described in section 4 and section 5 presents the results and predictive performance. Section 6 considers alternative approaches for measuring the impact of forecast accuracy on damages and its implications while section 7 concludes.

## 2 Modeling Framework

There are many potential determinants of hurricane damages. Natural hazards or forces such as the maximum wind speed of the hurricane, its minimum central pressure, the resulting storm surge and rainfall are all considered to be amongst the most important; see Nordhaus (2010), Murnane and Elsner (2012) and Chavas et al. (2017). Damages are also determined by the vulnerabilities associated with the population at a given location as measured by income, housing or capital stock; see Pielke Jr and Landsea (1998) and Pielke Jr et al. (2008). At the same time, damages can be mitigated through longer-term adaptation efforts associated with higher incomes, stricter building codes, levees and government spending on damage mitigation programs; see Bakkensen and Mendelsohn (2016), Geiger et al. (2016), Dehring and Halek (2013) and Davlasheridze et al. (2017).

I model hurricane damages using a Cobb-Douglas power function that is composed of natural forces  $\mathbf{F}_i$ , vulnerabilities  $\mathbf{V}_i$ , and adaptation  $\mathbf{A}_i$ . Taking logarithms gives the expression for damages from any given hurricane

$$\ln(d_i) = c + \alpha \ln(\mathbf{V}_i) + \beta \ln(\mathbf{F}_i) - \delta \ln\left(\mathbf{A}_i(\hat{\mathbf{F}}_i)\right), \quad (2.1)$$

where bold terms are vectors. This formulation is an extension of Bakkensen and Mendelsohn (2016) to allow for adaptation under imperfect information. I assume adaptation is described by the linear relationship from Bakkensen and Mendelsohn (2016):  $\ln(\mathbf{A}_i(\mathbf{F}_i)) = \boldsymbol{\gamma}_1 \ln(\mathbf{V}_i) + \boldsymbol{\gamma}_2 \ln(\mathbf{F}_i)$  where  $\boldsymbol{\gamma}_j$  are matrices and re-parameterize (2.1) by adding and subtracting adaptation under perfect information to get

$$\ln(d_i) = c + (\boldsymbol{\alpha} - \delta \boldsymbol{\gamma}_1) \ln(\mathbf{V}_i) + (\boldsymbol{\beta} - \delta \boldsymbol{\gamma}_2) \ln(\mathbf{F}_i) + \delta \boldsymbol{\gamma}_2 \left[ \ln(\mathbf{F}_i) - \ln(\hat{\mathbf{F}}_i) \right], \quad (2.2)$$

where the final term captures the distance between the actual and predicted natural forces. This allows me to study the implications of short-term accuracy for damages, which is the focus here, as well as longer-term accuracy from seasonal and climate predictions.

I extend the formulation further to capture the joint impact of forecast accuracy and uncertainty around the forecast. The final term in (2.2) can be replaced with a general measure of forecast uncertainty:

$$\ln(d_i) = c + (\boldsymbol{\alpha} - \boldsymbol{\eta} \boldsymbol{\gamma}_2^{-1} \boldsymbol{\gamma}_1) \ln(\mathbf{V}_i) + (\boldsymbol{\beta} - \boldsymbol{\eta}) \ln(\mathbf{F}_i) + \boldsymbol{\eta} \ln\left(U_i(\mathbf{F}_i, \hat{\mathbf{F}}_i)\right) + \epsilon_i, \quad (2.3)$$

where  $\boldsymbol{\eta} = \delta \boldsymbol{\gamma}_2$ ,  $U_i(\cdot)$  jointly captures accuracy and uncertainty for the forecast, and  $\epsilon_i$  indicates the residual or approximation error when going from (2.2) to (2.3).

This formulation embeds many existing models of hurricane damages. Emanuel (2005), Nordhaus (2010) and Strobl (2011) implicitly set  $\boldsymbol{\eta} \equiv \mathbf{0}$  and  $\boldsymbol{\alpha} \equiv \mathbf{1}$  to examine the relationship between damages and natural hazards. Others set  $\boldsymbol{\eta} \equiv \mathbf{0}$  to investigate the relationship between damages and vulnerabilities; see Kellenberg and Mobarak (2008) and Geiger et al. (2016). Bakkensen and Mendelsohn (2016) allow for  $\boldsymbol{\eta} \neq \mathbf{0}$  but implicitly assume  $U_i \equiv \mathbf{1}$  which means that they are unable to identify  $\boldsymbol{\eta}$  directly but have to do so indirectly.

There are many ways to measure forecast uncertainty. A popular measure is the mean square forecast error (MSE); see Ericsson (2001). Jurado et al. (2015) propose a time-varying MSE measure across variables using a dynamic factor model with stochastic volatility. Alternatively, the log score (see Mitchell and Wallis, 2011) evaluates the predicted density,  $\hat{g}_i(\cdot)$ ,

at  $F_i$  conditional on the prediction  $\hat{F}_i$ . When  $F_i$  falls in the tail of  $\hat{g}_i(\cdot)$ , it has a lower probability and so is associated with higher uncertainty. Another measure is the continuously ranked probability score, which compares against the cumulative distribution function.

Rossi and Sekhposyan (2015) propose a measure of forecast uncertainty based on the unconditional likelihood of the observed outcome. Their measure, in the context of a single variable, is computed by evaluating the predicted cumulative distribution function at  $F_{1i}$

$$U_{1,i}(F_{1i}, \hat{F}_{1i}) = \int_{-\infty}^{F_{1i}} \tilde{g}_i(x_1 | \hat{F}_{1i}) dx_1, \quad (2.4)$$

which captures how likely it is to observe  $F_{1i}$  given the predicted distribution. Their innovation is that  $\tilde{g}_i(\cdot)$  is computed using historical forecast errors. While Rossi and Sekhposyan (2015) focus on the full-sample distribution, it is possible to allow  $\tilde{g}_i(\cdot)$  to change across events (or time) so as to capture any changes in the predicted distribution; see Hendry and Mizon (2014). This measures the forecast accuracy in the context of the ex-ante uncertainty when the forecast was produced.

When  $\tilde{g}_i(\cdot)$  roughly follows a normal distribution then for small distances between  $F_{1i}$  and  $\hat{F}_{1i}$ , (2.4) can be approximated as

$$U_{1,i}(F_{1i}, \hat{F}_{1i}) \approx \frac{F_{1i} - \hat{F}_{1i}}{\hat{\sigma}_{i,(F_{1-}, \hat{F}_{1-})}}, \quad (2.5)$$

where  $\hat{\sigma}_{i,(F_{1-}, \hat{F}_{1-})}$  is the time-varying standard deviation of the predicted distribution based on historical forecast errors. It represents the ex-ante risk (in a Knightian sense) ascribed to the forecast at the time of the forecast. If (2.5) has an absolute value greater than one, then the forecast error falls outside of its expected mid-range and is associated with greater uncertainty. An absolute value less than one indicates there is less uncertainty since the forecast error is within the expected range. This is related to other comparisons of ex-post and ex-ante forecast uncertainty; see Clements (2014) and Rossi et al. (2017).

The measure is generalized further by taking logs and relaxing the fixed 1-to-1 relationship between forecast accuracy and ex-ante risk

$$\ln(U_{1,i}(F_{1i}, \hat{F}_{1i})) \approx \frac{\eta_1}{\eta} \ln(|F_{1i} - \hat{F}_{1i}|) + \frac{\eta_2}{\eta} \ln(\hat{\sigma}_{i,(F_{1-}, \hat{F}_{1-})}). \quad (2.6)$$

Plugging (2.6) back into the model for damages in (2.3) allows me to evaluate which aspect of forecast accuracy or uncertainty matters the most for hurricane damages.

### 3 Hurricane Forecast Errors, Uncertainty and Damages

The National Hurricane Center (NHC) maintains all hurricane forecasts that have been produced since it was established in 1954. The NHC’s ‘official’ forecasts form the basis for hurricane watches, warnings and evacuation orders and are widely distributed to and used by news outlets. The forecasts are not produced by a single model but are a combination of different models and forecaster judgment which changes over time; see Broad et al. (2007).

Forecasts of the track and intensity are generated every 6 hours for the entire history of a storm. While forecasts can extend out over 120-hours-ahead, I focus on the 12-hour-ahead track forecasts. This is motivated by the fact that individuals typically wait until the last minute for their adaptation efforts and often focus on the forecast track; see U.S. Army Corps of Engineers (2004) and Milch et al. (2018). The track is also an integral part of the NHC’s forecasts of intensity including rainfall (Kidder et al., 2005), wind speed (DeMaria et al., 2009), and storm surge (Resio et al., 2017). The 12-hour-ahead forecasts are available for virtually every U.S. hurricane going back to 1955.<sup>1</sup>

I start by computing the 12-hour-ahead forecast errors from every tropical storm in the North Atlantic since 1954. This includes 14,641 forecast errors from 744 storms. Forecast errors are calculated as the distance between the actual track of the storm using Vincenty (1975)’s formula for the surface of a spheroid. Next, I estimate the time-varying forecast error densities for every year since 1955 using a rolling window of the past five years.<sup>2</sup>

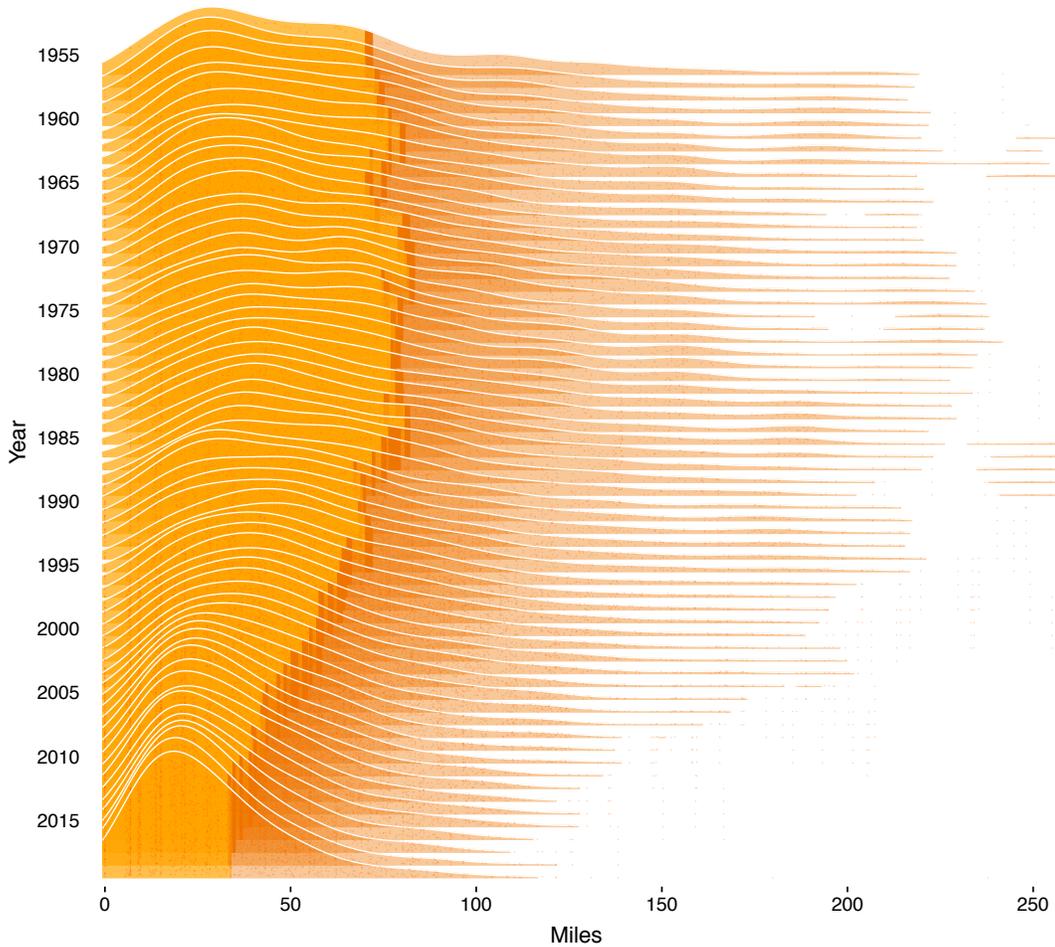
Figure 3.1 shows how the densities have changed over time. Historically they were skewed to the right with long fat tails due to large outliers. Skewness declined dramatically after 1985 and a truncated normal distribution now appears to be a good approximation. Improved accuracy was driven by the use of satellites and supercomputers; see Rappaport et al. (2009).

The 67th percentile of each density is similar to how the NHC measures the radius of the ‘cone of uncertainty’. The radius is measured as the distance that captures 2/3 of all forecast errors in the past five years and denotes the uncertainty associated with the forecast location.<sup>3</sup> Thus, I refer to this as the time-varying ex-ante uncertainty.

<sup>1</sup>No forecasts are available for Debra [1959] and Ethel [1960], likely due to their short duration. Longer horizon forecasts are available for even fewer storms and intensity forecasts are only available since 1990.

<sup>2</sup>Estimates prior to 1959 are based on samples less than five years since the database only goes back to 1954.

<sup>3</sup>NHC used to measure the radius based on the previous ten years; see Broad et al. (2007).

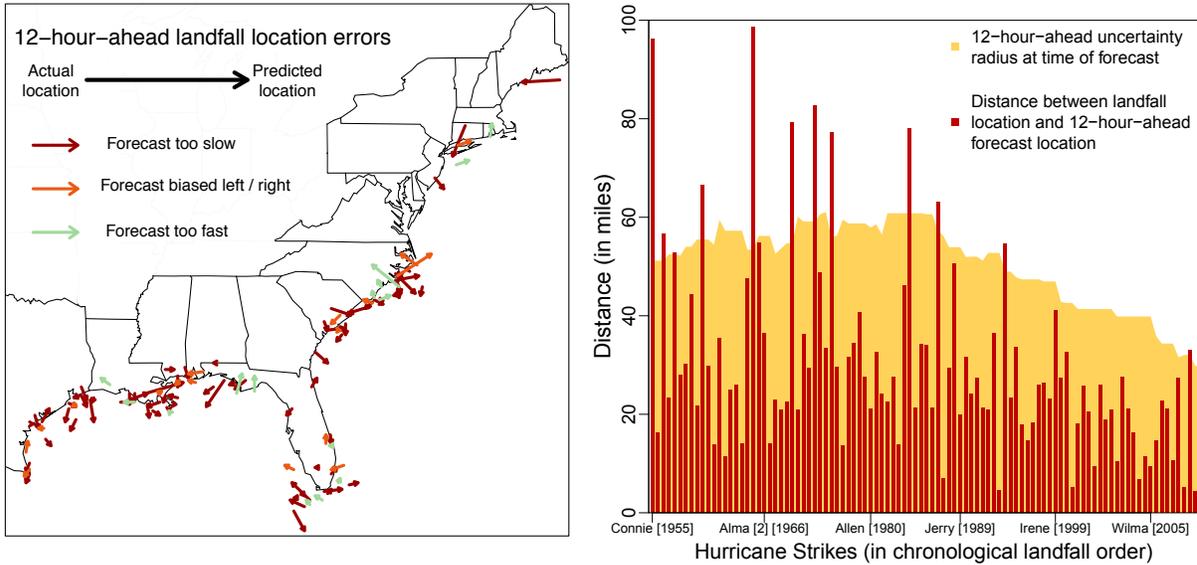


Notes: Densities are estimated using all 12-hour forecast errors for all storms in the Atlantic Basin over the previous five years. This amounts to 100-2000 observations per density. The dark orange bars represent the 67th percentile in each year.

**Figure 3.1: Atlantic Basin Tropical Storm Forecast Error Densities 1955-2019**

The most relevant forecasts for hurricane damages are those that are made just before landfall. In total, 88 hurricanes made landfall between 1955 and 2015. Accounting for the fact that some hurricanes struck multiple locations, i.e. Katrina [2005] first crossed the Florida panhandle and then moved into the Gulf and struck Louisiana several days later, there were 101 strikes. I focus on the 98 strikes for which forecasts are available.

Since the forecast and location of the hurricane is only updated at six-hour intervals, I round the timing of each landfall to the closest point in that interval. So, a hurricane landfall at 16:00 Universal Time (UTC) is rounded to 18:00 UTC. I subtract the length of the forecast horizon from the landfall time to get the time at which the 12-hour-ahead ‘landfall forecast’ was generated. So the 12-hour-ahead landfall forecast of a storm that made landfall at 18:00 UTC was generated at 6:00 UTC.



Panel A: Direction of 12-hour-ahead landfall errors in space

Panel B: 12-hour-ahead landfall errors over time

Notes: The 12-hour-ahead landfall errors are computed such that the forecast was made 12 hours before the closest observed point of the hurricane at landfall. In panel A, the coloring of arrows is based on the calculated angle of the triangle which is calculated from the distance between the forecast and the actual (shown), the distance between the actual 12-hour prior and the current actual and the distance between the previously observed value and the forecast. Green is less than 90 degrees, orange is greater than 90 but less than 135 degrees and red is greater than 135 degrees. In panel B, the shaded area is the ex-ante uncertainty and corresponds to the distance that captures 2/3 of all forecast errors in the five years prior to the year in which the strike occurred.

**Figure 3.2: 12-Hour-Ahead Hurricane Landfall Errors**

I then compute the landfall forecast error by calculating the distance between the forecast made 12-hours before landfall and the location at landfall. Although the forecast errors are purely distance measures, it is possible to derive forecast error directions based on the location of the hurricane when the forecast was generated. Panel A of Figure 3.2 plots the spatial distribution of the landfall locations and forecast error directions for most hurricane strikes since 1955. It shows that there is a wide geographic distribution along the Gulf and Atlantic coasts with landfalls ranging from southern Texas to eastern Maine. There also does not appear to be any systematic pattern in the error directions.

Panel B of Figure 3.2 plots the 12-hour-ahead landfall forecast errors for all hurricane strikes along with the estimated ex-ante uncertainty from 1955-2015. The large variation in the forecast errors over time will allow me to identify if they have any impact on damages. Although ex-ante uncertainty has declined gradually since the 1990's, the forecast errors themselves do not have a clear trend despite the fact that larger errors do occur somewhat more often earlier on in the sample. Forecast errors exceed the ex-ante uncertainty in about 12 percent of all strikes; most recently for Sandy [2012].

I collate hurricane strike damages from multiple sources. I primarily rely on annual Atlantic Hurricane Season reports following Pielke Jr and Landsea (1998).<sup>4</sup> I supplement and update these numbers using the latest tropical cyclone report for each storm and Blake et al. (2011). The resulting damages are similar to Pielke Jr et al. (2008). However, they are considerably higher than NOAA’s ‘Storm Events’ database (which suffers from under-reporting; see Smith and Katz, 2013) and are somewhat lower than NOAA’s ‘Billion-Dollar’ database, which has a broader definition of damages; see Weinkle et al. (2018).<sup>5</sup>

## 4 Model Selection Methods

While the relationship between hurricane damages and forecast accuracy is the primary interest of the analysis, there are many other potential determinants of damages. I use model selection to focus the analysis on those variables which are most important for damages. The approach I follow is broadly defined as an automatic general-to-specific (*Gets*) modeling framework, which is described in detail by Hendry and Doornik (2014). Developments by Hendry et al. (2008), Castle et al. (2012), Castle et al. (2015), Hendry and Johansen (2015) and Pretis et al. (2016), illustrate its usefulness across a range of applications.

The approach starts with the formulation of a general unrestricted model (GUM) that includes all potentially relevant explanatory variables. It is assumed that the residuals of the GUM are iid normal but this can be relaxed in extensions of the framework

$$y_i = \sum_{s=1}^S \beta_s x_{s,i} + \sum_{n=1}^N \gamma_n z_{n,i} + \epsilon_i \quad \epsilon_i \sim \text{IN} [0, \sigma_\epsilon^2] \quad \{i = 1 \dots J\}. \quad (4.1)$$

Theory-driven variables that are not selected over are denoted by  $x_s$  and all variables that will be selected over are denoted by  $z_n$ . It is also assumed that the GUM is potentially sparse and nests the underlying local data generating process (LDGP). This ensures that *Gets* consistently recovers the same model as if selection began from the LDGP. Thus, formulation of the GUM is an integral part of the modeling process.

The properties of *Gets* model selection are most easily illustrated by considering the case where each  $z_n$  is orthogonal. In this case, selection can be performed by computing the

<sup>4</sup>Reports were published in the Monthly Weather Review through 2011 and are available from the Hurricane Research Division until 2011. The National Hurricane Center maintains the annual summaries since 2012.

<sup>5</sup>The Storm Events database forms the basis for the SHELDUS database which is available since 1959. Despite its disaggregation of damages across counties, the SHELDUS database tends to underreport damages and often allocates them equally across counties.

sample of t-statistics under the null hypothesis that  $H_0 : \gamma_n = 0 \forall n \in N$ . The t-statistics can then be ordered from largest to smallest

$$t_{(N)}^2 \geq \dots \geq t_{(m)}^2 \geq c_a^2 \geq t_{(m-1)}^2 \geq \dots \geq t_{(1)}^2, \quad (4.2)$$

where  $\alpha$  determines the critical value above which variables are retained. In this setup the decision to select is only made once, i.e. ‘one-cut’. Under the null hypothesis, the average number of irrelevant variables retained is

$$k = \sum_{i=0}^N i \frac{N!}{i! (N-i)!} \alpha^i (1-\alpha)^{N-i} = N\alpha, \quad (4.3)$$

which shows that  $\alpha$  determines the false-retention rate of variables in expectation, i.e. the ‘gauge’ (Castle et al., 2011 and Johansen and Nielsen, 2016). The gauge plays the same role as regularization does in other machine learning or model selection procedures in that it controls the loss of information in the selection procedure; see Mullainathan and Spiess (2017). However, unlike other regularization parameters which are determined based on in-sample model performance, the gauge has a theoretical interpretation and is typically chosen based on the size of the GUM,  $\alpha = 1/N$ , so that *on average* one irrelevant variable is kept.

If the variables are correlated then ‘one-cut’ selection is no longer appropriate. However, the approach can be extended by sampling different blocks of variables at a time and ensuring that any reduction satisfies the ‘encompassing principle’ as determined by the gauge; see Doornik (2008). This can be augmented with a multi-path tree search and diagnostic tests so as to minimize potential path dependencies and to ensure a statistically well-specified model; see Hendry and Doornik (2014).

The final model provides a parsimonious explanation of the GUM conditional on the acceptable amount of information loss as determined by the gauge. When multiple models are retained then information criteria are used to select between otherwise equally valid models. Alternatively, ‘thick modeling’, as proposed by Granger and Jeon (2004) and discussed in a *Gets* framework by Castle (2017), can be used to pool selected models.

There are many different model selection algorithms available. I use the multi-path block search algorithm known as ‘Autometrics’ available in PcGive; see Doornik (2009) and Doornik and Hendry (2013). An alternative multi-path search algorithm is implemented using the ‘gets’ package in R; see Pretis et al. (2018). I also compare the results by performing

model selection using regularized regression methods (i.e. Lasso) as implemented in the ‘glmnet’ package in R; see Friedman et al. (2010).

## 5 Selecting a Model of Damages

The GUM that I estimate includes most potential determinants of hurricane damages and several controls for spatial and temporal heterogeneity. It contains 37 explanatory variables and is estimated over a sample of 98 observations. Lower case variables are in logs:

$$\begin{aligned} \text{damage}_i = & c + \alpha_1 \text{hd}_i + \alpha_2 \text{ih}_i + \alpha_3 \text{FREQ}_i + \beta_1 \text{rain}_i + \beta_2 \text{surge}_i + \beta_3 \text{press}_i \\ & + \beta_4 \text{wind}_i + \beta_5 \text{MOIST}_i + \beta_6 \text{ace}_i + \beta_7 \text{GST}_i + \eta_1 \text{forc12}_i + \eta_2 \text{radii12}_i \\ & + \delta_1 \text{STREND}_i + \delta_2 \text{YTREND}_i + \psi \text{MONTH}_i + \kappa \text{HOUR}_i + \lambda \text{STATE}_i + \epsilon_i. \end{aligned} \quad (5.1)$$

The first line includes the vulnerabilities (V): housing unit density (hd), income per housing unit (ih) and ‘real-time’ hurricane strike location frequency (FREQ). The first two lines list the natural hazards (F): maximum rainfall (rain), storm surge (surge), negative minimum central pressure (press), maximum wind speed (wind), soil moisture relative to trend (MOIST), accumulated cyclone energy (ace) and global surface temperature (GST).<sup>6</sup> The second line captures forecast accuracy and uncertainty (U): 12-hour-ahead forecast track errors (forc12) and the ex-ante uncertainty (radii12). The last line lists additional spatial and temporal controls: strike and annual trends, month dummies, hour dummies (to control for the six-hour period in which the landfall occurred), and U.S. state dummy variables. See Appendix B for further details on how the additional variables were constructed.

I use *Gets* model selection to discover the most important drivers of hurricane damages without selecting over forecast accuracy and uncertainty. Since the GUM has 35 variables to select over, I set the target gauge equal to  $\frac{1}{35} \approx 0.03$ . There are  $2^{35}$  ( $> 34$  billion) possible model combinations when allowing for every variable to be selected over. For a target gauge of 3 percent, the selection algorithm narrows the search space to  $2^{17}$  ( $< 200$  thousand). It eliminates entire branches of models and ultimately only estimates 304 candidate models. The algorithm finds 8 terminal models as acceptable reductions of the GUM; see Appendix Table A.1. The final model is selected using the Bayesian information criterion (BIC).

Column (1) of Table 5.1 illustrates the most important drivers of damages using *Gets*. Wind speed does not appear. Instead, minimum central pressure is included. Even when

<sup>6</sup>Radiative forcing could also be included as in Morana and Sbrana (2019).

wind speed is included in the model with central pressure, see Appendix Table A.1, the estimated coefficient is the wrong sign. This is consistent with Bakkenen and Mendelsohn (2016), who find that central pressure provides a more reliable explanation of damages. I find more disaggregated measures of hazards with rainfall and storm surges: rainfall causes inland flooding whereas storm surges damage the coast. I also find that greater housing density and higher incomes are associated with more damage.

**Table 5.1: Selected Models of Hurricane Damages**

Selection Target:	(1) Gets	(2) Lasso	(3) Robust Gets
	3%	BIC (9%)	1%
Housing density	0.40*** (0.15)	0.40*** (0.15)	0.24** (0.09)
Income per housing unit	1.30*** (0.20)	1.30*** (0.20)	1.76*** (0.19)
Historical frequency		-3.38 (2.20)	
Max rainfall	1.15*** (0.31)	0.96*** (0.32)	0.56*** (0.20)
Max storm surge	1.34*** (0.38)	1.36*** (0.37)	0.98*** (0.24)
Min central pressure (-)	52.2*** (8.58)	50.2*** (8.45)	57.8*** (5.41)
Seasonal cyclone energy		0.57** (0.24)	
Forecast errors	0.47* (0.24)	0.48** (0.24)	0.27* (0.15)
Ex-ante uncertainty	0.17 (0.98)	0.78 (1.00)	2.10** (0.96)
Outliers:	No	No	Yes
Nonlinear terms:	No	No	Yes
$\hat{\sigma}$	1.268	1.233	0.795
$R^2$	0.821	0.834	0.930
BIC	3.673	3.688	2.774

*Notes: Estimated using 98 observations including a constant and dummy variables for Gerda [1969] and Floyd [1987]. Standard errors are in parentheses. \* $p < 0.1$  \*\* $p < 0.05$  \*\*\* $p < 0.01$*

The results change with other selection methods. Lasso shrinks the estimated coefficients using a penalty, which is chosen here using BIC. The post-selection OLS coefficients (and standard errors) are shown in column (2). All of the variables retained by Gets are also retained by Lasso. However, an additional hazard and a proxy for long-term adaptation are both kept, which is consistent with the fact that BIC corresponds to a looser target gauge

of 9 percent for the given sample size and number of variables; see Campos et al. (2003).

To account for possible model misspecification and to assess the robustness of the selected model, I extend the *Gets* model by selecting over strike fixed effects to capture outliers relative to the model using impulse indicator saturation (IIS; see Hendry et al., 2008 and Johansen and Nielsen, 2016) and over squares of the explanatory variables to capture potential non-linearities. I find a non-linear relationship between income and damages; which is consistent with Geiger et al. (2016). I also find multiple outlying strikes which capture measurement issues in the 1950's as well as glancing landfalls; see Appendix Table A.2. Since the outliers are all negative and roughly the same magnitude, I combine them into a single dummy variable to obtain a more parsimonious model presented in column (3).

Forecast accuracy is significant across the different models. In each case an increase in the errors is associated with an increase in damages even after controlling for ex-ante uncertainty. To put their relative importance into context, the standardized coefficients from the Robust Gets model indicate that it takes roughly a 5 standard deviation shock for forecast errors to have the same impact on hurricane damages as a 1 standard deviation shock to central pressure. This is roughly the same size that is required for rainfall. A simple back of the envelope calculation indicates that, on average, a one standard deviation reduction in the distance between where a hurricane is expected to strike and where it actually strikes is associated with a decline in damages by up to \$9,000 in per affected household. This represents about 18 percent of the typical insurance claim or about 4 percent of the total replacement value of a typical at risk home in 2018.<sup>7</sup>

Next I consider how sensitive the results are to alternative measures of forecast accuracy and uncertainty. Table 5.2 presents the results. Column (1) presents a restricted robust gets model which excludes ex-ante uncertainty. Column (2) presents a theory consistent restriction of the robust gets model by imposing a fixed relationship between the forecast errors and the radius of uncertainty; see (2.5). In both cases these restrictions are strongly rejected. Column (3) uses Rossi and Sekhposyan (2015)'s measure of uncertainty in (2.4) with a non-parametric time-varying forecast-error distribution based on rolling window of the past five years of forecast errors. Column (4) considers a weighted RMSE measure.

---

<sup>7</sup>Based on estimates from CoreLogic and FEMA.

The results indicate that there is a positive and statistically significant link between forecast accuracy and damages across each of the measures. The size of this relationship is fairly stable regardless of which measure of accuracy or uncertainty is considered. Columns (3) and (4) indicate that imposing a normal distribution on the historical forecast errors does not substantially alter the parameter estimate or increase the noise.

**Table 5.2: Sensitivity to Measures of Accuracy and Uncertainty**

	(1) Errors	(2) Errors/Radius	(3) R&S(2015)	(4) RMSE
Housing density	0.27*** (0.10)	0.28*** (0.10)	0.28*** (0.10)	0.26*** (0.10)
Income per housing unit	1.43*** (0.12)	1.36*** (0.12)	1.36*** (0.12)	1.44*** (0.13)
Min central pressure (-)	56.6*** (5.49)	56.8*** (5.55)	56.9*** (5.58)	57.3*** (5.51)
Max rainfall	0.57*** (0.21)	0.55*** (0.21)	0.54** (0.21)	0.56*** (0.21)
Max storm surge	0.99*** (0.24)	0.97*** (0.25)	0.97*** (0.25)	0.95*** (0.24)
Forecast errors	0.34** (0.15)			
Errors/ex-ante		0.29* (0.16)		
Uncertainty			0.23* (0.14)	
Weighted RMSE				0.37* (0.19)
Outliers:	Yes	Yes	Yes	Yes
Nonlinear terms:	Yes	Yes	Yes	Yes
$\hat{\sigma}$	0.812	0.819	0.822	0.817
$R^2$	0.927	0.925	0.925	0.926
BIC	2.781	2.799	2.806	2.795

*Notes: All equations are estimated using 98 observations and include a constant and a dummy variable for Gerda [1969] and Floyd [1987]. Standard errors are in parentheses. R&S(2015): Rossi and Sekhposyan (2015). RMSE is weighted using an AR(1) coefficient of 0.3 on forecast errors after the initial strike. \* $p < 0.1$  \*\* $p < 0.05$  \*\*\* $p < 0.01$*

To assess the stability of the models and to ensure that they are not overfitting the data, I evaluate the out-of-sample performance for hurricane strikes between 2016 and 2019. I assess performance relative to two theory-driven models; see Nordhaus (2010) and Strobl (2011) and Bakkensen and Mendelsohn (2016). These models are much simpler in that they only include one natural hazard (central pressure) and one vulnerability (income). To ensure that outliers do not bias the results, I estimate each of these models with a glancing strike

**Table 5.3: Relative Out-of-Sample Model Performance**

	(1) Bias	(2) RMSE	(3) MAPE
Gets	0.91	0.75	0.80
Lasso OLS	0.95	0.77	0.94
Robust Gets	<b>0.20</b>	<b>0.20</b>	<b>0.64</b>
Nordhaus (2010)	1.07	0.99	5.46

*RMSE: Root mean square error. MAPE: Mean absolute percent error. Metrics are computed relative to the performance of the Bakkenen and Mendelsohn (2016) model. Values smaller than 1 indicate better performance.*

dummy variable and without ex-ante uncertainty due to its general insignificance. This ensures that no model has an undue advantage. Thus, the only difference between the Gets model and the Robust Gets model is that the latter includes a non-linear income term.

Table 5.3 shows that each of the selected models outperforms the theory-driven models across each of the accuracy metrics. However, the robust model does substantially better than any other model. This is driven by Harvey [2017] where the robust model almost perfectly predicts official damages. However, as Appendix Figure A.1 shows, the robust model consistently outperforms the other models. This suggests that performance is substantially improved when accounting for the non-linear effect of income. It also shows that the models are robust without overfitting the data.

## 6 Measuring the Impact of Forecast Accuracy

I now consider three different approaches for identifying the impact of forecast accuracy on hurricane damages. The first considers inference in the context of model selection and confirms the above findings. The second treats the selected model as a possibly misspecified structural equation and uses instruments to address concerns of omitted variable bias and endogeneity. The third approach tests for the existence of a causal relationship using model invariance and allows me to perform counterfactual exercises. Together these approaches provide a holistic view of the impact of forecast accuracy on hurricane damages.

### 6.1 Post-Selection Inference

While model selection is typically used for prediction, it can also be used for inference. However, when conducting inference on selected variables it is important to account for uncertainty in the selection procedure by adjusting the standard errors or critical values;

see Berk et al. (2013) and Van de Geer et al. (2014). Alternatively, it is possible to restrict selection such that the variables of interest for inference are always retained and so does not require other adjustments; see Chernozhukov et al. (2018). The results in Table 5.1 implicitly follow this latter approach since neither forecast errors nor ex-ante uncertainty were selected over. However, the results do not account for correlation between the variables forced in the model and those that were selected over. Therefore, I consider two procedures that control for these correlations and allow for valid inference.

The first approach is to orthogonalize the variables that will be selected over by the variables that will not. This procedure was proposed by Hendry and Johansen (2015) to test the significance of theory relevant variables after performing model selection. I orthogonalize all other variables by forecast accuracy, ex-ante uncertainty and income, which ensures that the relationship between income and damages is not falsely attributed to the relationship between forecast accuracy and damages.

The results are presented in column (1) of Table 6.1. The selected variables are identical to the previous Gets result in Table 5.1, however the coefficients on forecast errors and ex-ante uncertainty are different. The coefficient for forecast errors is slightly smaller but remains statistically significant at the 10% level. On the other hand, the coefficient on ex-ante uncertainty is now negative but is not statistically significant from zero. This illustrates that the correlations between variables play an important role when interpreting the results for ex-ante uncertainty.

The second approach is a double Lasso procedure whereby variables that explain forecast accuracy or uncertainty are retained regardless of whether they matter for damages. This was proposed by Belloni et al. (2014) in the context of treatment variables in order to protect against omitted variable bias due to selection. The procedure is described as follows: Run a Lasso regression on the full model of damages excluding forecast accuracy and ex-ante uncertainty. Next, run a Lasso regression for the models of forecast accuracy and ex-ante uncertainty using all other explanatory variables. Finally, combine the variables selected in the first two stages along with forecast errors and ex-ante uncertainty into an OLS regression on damages. This allows for valid inference on the forecast errors and ex-ante uncertainty when they are conditionally exogenous.

**Table 6.1: The Impact of Forecast Accuracy on Damages**

	(1) Orthogonal <i>Gets</i>	(2) Double Lasso	(3) IV
Selection Target:	3%	BIC (9%)	N/A
Forecast errors	0.39* (0.23)	0.50** (0.21)	0.49*** (0.17)
Ex-ante Uncertainty	-1.58 (1.13)	0.98 (1.01)	2.95** (1.25)
Obs.	98	98	65

*Notes: HAC standard errors are in parentheses: \* $p < 0.1$  \*\* $p < 0.05$  \*\*\* $p < 0.01$*

The results for these two coefficients are shown in column (2) of Table 6.1. The coefficient for forecast errors is different from zero at a 5% significance level and is broadly consistent with the results in column (1) and those in Table 5.1. On the other hand, the coefficient on ex-ante uncertainty is positive and statistically insignificant.

## 6.2 Exogenous Instruments

The post-selection inference approaches assume that forecast errors and ex-ante uncertainty are exogenous conditional on the variables in the GUM. If this assumption does not hold then the procedures will produce biased estimates of the true parameters. Therefore it is important to include additional variables in the model or to use of an exogenous instrument to ensure this assumption is not violated.

There are several reasons why the forecast errors may not be exogenous. Errors may be correlated with storm dynamics where a volatile storm that is more difficult to forecast causes more damage. For example, the rapid intensification of a hurricane prior to landfall is difficult to forecast and is associated with higher damages; see Kaplan et al. (2010). Errors may be correlated with long-term adaptation and technological change. This is especially true when considering that the measure of ex-ante uncertainty is effectively a five-year moving percentile of forecast errors and so is driven more by longer-term trends such as government expenditures and technological change. Both of these effects could bias the estimated relationship between forecast accuracy and damages.

I construct an instrument based on a measure of forecast skill in order to address concerns about storm dynamics. Forecast skill is often used to assess the performance of hurricane and weather forecasts over time; see Cangialosi and Franklin (2016). I calculate forecast

skill as the inverse ratio of the official forecast errors to naïve forecast errors from a simple climatology and persistence model (available since 1970). As a result, forecasts for an erratic storm that is more difficult to forecast will not be penalized with lower skill. Assuming that both types of forecast errors are log normal and affected in the same way by the endogenous storm dynamics, then the log forecast skill is a valid instrument by removing the endogeneity.

I also include several additional controls. To control for longer-term adaptation efforts I include the normalized length of protective levees from the US Army Corps of Engineers' National Levee Database as a measure of how location-specific adaptation efforts have changed over time. To control for the role of hurricane evacuations I use hurricane warning lead times. Earlier warnings are given for stronger storms or when they are expected to strike more populated locations. Earlier warning times can also cover larger areas due to the uncertainty associated with the path of the hurricane. I address this by looking at the number of hours in advance a warning was given to the eventual landfall location divided by the length of coastline under warning.

I perform two-stage least squares starting from the robust Gets model with the additional controls. Forecast skill is highly significant in the first stage regression with a t-value of  $-7.4$ . Table 6.1 shows the second stage estimates for the instrumented forecast errors and ex-ante uncertainty. The coefficient on instrumented forecast errors is highly significant at a 1% level and is similar in magnitude to the other estimates despite being based on a smaller sample starting in 1970. Ex-ante uncertainty remains positive and is different from zero at a 5% significance level. Thus, the results from this exercise suggest that the findings are robust to additional controls and accounting for endogeneity concerns.

### **6.3 Model Invariance and Valuing Improved Forecast Accuracy**

Another approach is to identify causal relationships through model invariance; see Engle et al. (1983), Peters et al. (2016), and Arjovsky et al. (2019). The idea behind this is that if historical shocks to forecast errors only impact damages through forecast errors, then the model of damages is invariant to changes in the forecast errors and there is a likely causal relationship between forecast errors and damages; see Hendry (1995, Chapter 5). Several tests of model invariance have been proposed in the time-series literature by Engle and Hendry (1993), Hendry and Santos (2010), and Castle et al. (2017).

**Table 6.2: A Test of Model Invariance**

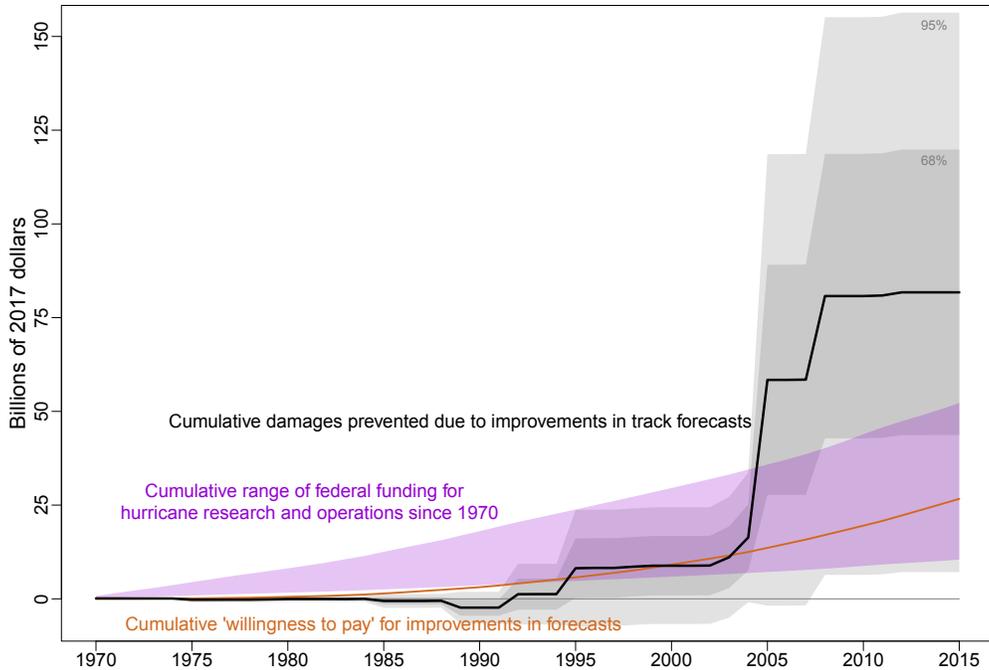
	(1) Marginal + IIS	(2) Augmented Conditional
Retained Impulses:	Forecast Errors	Damages
Connie [1955]	1.29 (0.57)**	-0.96 (0.96)
Cleo [1964]	-1.21 (0.51)**	1.38 (0.87)
Opal [1995]	-1.48 (0.52)***	0.24 (0.89)
Isabel [2003]	-1.19 (0.50)**	0.98 (0.85)
Arthur [2014]	-1.73 (0.54)***	-0.53 (0.94)
F-statistic	.	1.05
p-value	.	[0.394]

Notes: Regressions are estimated using 98 observations with all regressors from the Robust Gets model. IIS is run with a 1% gauge. Standard errors are in parentheses. The tail probability associated with the null hypothesis of the F-statistic with (5,83) degrees of freedom is in square brackets. \*p< 0.1 \*\*p< 0.05 \*\*\*p< 0.01

To test the invariance of the robust Gets model with respect to the forecast errors, I follow the approach in Hendry and Santos (2010). First, I run impulse indicator saturation (IIS) on the marginal model of forecast errors regressed on all other explanatory variables. I retain any impulses detected in this model. Next, I estimate the conditional model of damages regressed on the forecast errors, all other explanatory variables, and the retained impulses from the marginal model. Finally, I test the joint significance of these impulses in the conditional model. The null hypothesis is that the model is invariant so that any shocks to the marginal model pass directly through the forecast errors to damages and do not otherwise change the parameters of the conditional model.

Table 6.2 shows that five outlying storms are detected in the marginal equation for forecast errors: Connie [1955], Cleo [1964], Opal [1995], Isabel [2003] and Arthur [2014]. However, after adding them to the conditional model and testing for their joint significance, a test statistic of 1.05 is obtained which when compared against an  $F$ -distribution with (5,83) degrees of freedom indicates that the null hypothesis of model invariance is not rejected with a p-value of 39.4 percent. This provides strong evidence in support of a causal relationship between the forecast errors and hurricane damages in this model.

The robust Gets model can also be used to conduct counterfactual experiments to assess the short-term impact of changes in forecast accuracy on hurricane damages. This requires that the parameter linking forecast accuracy to hurricane damages is identified such that model is semi-structural. The relative stability of the parameter estimate across different



Notes: Calculated as the difference in damages predicted from the Robust Gets model using the actual forecast error vs. damages predicted using the average forecast error from 1955-1969. The 5% - 95% confidence interval around this estimate is computed using the delta method assuming independence and normality. Federal funding for hurricane research and operations comes from past editions of the Office of the Federal Coordinator for Meteorology's Federal Plan for Meteorological Services and Supporting Research and is between 7% and 33% of total funding for meteorological operations and research costs following National Science Board (2007, see footnote 46). Willingness to pay is calculated as the cumulative sum of the population in every coastal county struck by a hurricane since 1970 times the real value of \$14.73 over time.

**Figure 6.1: The cost and benefits of improving forecast accuracy since 1970**

approaches and the failure to reject model invariance when there are large shocks to the forecast errors all suggest that the true parameter is identified. Assuming this is true, then the robust Gets model can be used to predict damages conditional on the forecast error for each strike since 1970 being equal to the average landfall-forecast error from 1955-1969. The model can also be used to predict damages conditional on the actual forecast error. The difference between these predictions is a prediction of the damages prevented due to forecast improvements since 1970. Under additional independence and normality assumptions it is possible to construct a 95 percent point-wise confidence interval around this prediction.<sup>8</sup>

Figure 6.1 shows that the predicted cumulative damages prevented due to forecast improvements since 1970 was around \$82 billion in 2015. Putting this in context, \$82 billion is 14 percent of the total predicted damages from 1970-2015 and is just shy of the total damage caused by Maria [2018]. This is driven both by improvements in forecast accuracy and by more damaging storms. Damages prevented are most pronounced for the 2005 and 2008

<sup>8</sup>See Technical Appendix C for details.

hurricane seasons which saw several very destructive hurricanes and above average forecast accuracy. On the other hand, despite the destructiveness of Sandy [2012], there was little damage prevented as the forecasts were not substantially different from the pre-1970 average.

To assess the relative benefits, I compare the prediction against two different estimates of costs. The first is the range of cumulative costs of producing the forecasts and their improvements since 1970.<sup>9</sup> After accounting for these costs, the predicted net savings is around \$30 – 71 billion. The second measure of cost is the cumulative private willingness to pay for forecast improvements since 1970 as extrapolated from Katz and Lazo (2011). The benefits of forecast improvements are greater than both public spending and private willingness to pay for them. This suggests that both individuals and the U.S. federal government underestimate the value of improving hurricane forecasts. Note that this estimate should be considered a lower bound of the total net benefits from forecast improvements since fatalities prevented (Willoughby et al., 2007) and reduced evacuation and damage mitigation costs (Regnier, 2008) are not included.

## 7 Conclusions

In this paper I evaluate the relationship between forecast accuracy and hurricane damages using a model of damages for all hurricanes to strike the continental United States in the past 60 years. I start with many possible drivers of damages and use model selection methods to determine the most important ones. I show that a small sub-set of drivers explains most of the historical damages and performs well for the latest hurricane strikes.

Despite specifying a richer model than previous studies, I find that forecast accuracy matters for damages. This relationship is positive, statistically significant and is robust to outliers, alternative measures of accuracy, model specifications, out-of-sample storms, and additional controls. On average, a one standard deviation increase in the distance between the storm’s predicted and actual landfall location translates on average into roughly \$9,000 additional dollars in damages per affected household. This value is equivalent to about 18 percent of the typical flood insurance claim or 4 percent of the total replacement value of a typical home at risk for hurricane damage in 2018.

---

<sup>9</sup>Obtained from the Office of the Federal Coordinator for Meteorology.

Using a counterfactual exercise, I show that improvements in the forecasts since 1970 have resulted in total damages being approximately \$82 billion less than they otherwise would have been. Although damages increased due to changes in vulnerabilities and natural hazards, improvements in forecast accuracy along with longer-term adaptation efforts have kept damages from rising faster than they otherwise would have. I find that there is a net benefit of around \$30 – 71 billion when compared against the realized cost of producing the forecasts. This illustrates that improvements in the forecasts produce benefits beyond the well-documented reduction in fatalities and have outweighed the associated costs.

This is particularly important since hurricanes are expected to become more difficult to forecast in the future. Knutson et al. (2010) argue that climate change will increase hurricane intensity and there will be more hurricanes like Harvey [2017] whose dynamics are hard to predict (Emanuel, 2017a,b). In light of this reality these findings support maintained investment in and further measures to improve hurricane forecasting capabilities along with other longer-term adaptation efforts so that any future loss of life and property is minimal.

## Bibliography

- Arjovsky, M., L. Bottou, I. Gulrajani, and D. Lopez-Paz (2019). Invariant risk minimization. *arXiv preprint arXiv:1907.02893*.
- Bakkensen, L. A. and R. O. Mendelsohn (2016). Risk and adaptation: evidence from global hurricane damages and fatalities. *Journal of the Association of Environmental and Resource Economists* 3(3), 555–587.
- Beatty, T. K., J. P. Shimshack, and R. J. Volpe (2019). Disaster preparedness and disaster response: Evidence from sales of emergency supplies before and after hurricanes. *Journal of the Association of Environmental and Resource Economists* 6(4), 633–668.
- Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies* 81(2), 608–650.
- Berk, R., L. Brown, A. Buja, K. Zhang, L. Zhao, et al. (2013). Valid post-selection inference. *The Annals of Statistics* 41(2), 802–837.
- Blake, E. S., E. N. Rappaport, and C. W. Landsea (2011). *The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)*. NOAA, National Weather Service, National Centers for Environmental Prediction, National Hurricane Center.
- Broad, K., A. Leiserowitz, J. Weinkle, and M. Steketee (2007). Misinterpretations of the "Cone of Uncertainty" in Florida during the 2004 hurricane season. *Bulletin of the American Meteorological Society* 88(5), 651–667.
- Campos, J., D. F. Hendry, and H.-M. Krolzig (2003). Consistent model selection by an automatic gets approach. *Oxford Bulletin of Economics and Statistics* 65, 803–819.
- Cangialosi, J. P. and J. L. Franklin (2016). 2015 Hurricane Season. NHC Forecast Verification Report, NOAA: National Hurricane Center.
- Castle, J. L. (2017). Sir Clive WJ Granger Model Selection. *European Journal of Pure and Applied Mathematics* 10(1), 133–156.
- Castle, J. L., J. A. Doornik, and D. F. Hendry (2011). Evaluating automatic model selection. *Journal of Time Series Econometrics* 3(1).
- Castle, J. L., J. A. Doornik, and D. F. Hendry (2012). Model selection when there are multiple breaks. *Journal of Econometrics* 169(2), 239–246.
- Castle, J. L., J. A. Doornik, D. F. Hendry, and F. Pretis (2015). Detecting location shifts during model selection by step-indicator saturation. *Econometrics* 3(2), 240–264.
- Castle, J. L., D. F. Hendry, and A. B. Martinez (2017). Evaluating forecasts, narratives and policy using a test of invariance. *Econometrics* 5(3), 39–65.

- Chavas, D. R., K. A. Reed, and J. A. Knaff (2017). Physical understanding of the tropical cyclone wind-pressure relationship. *Nature Communications* 8(1360).
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duffo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21(1), C1–C68.
- Clements, M. P. (2014). Forecast uncertainty—Ex Ante and Ex Post: U.S. inflation and output growth. *Journal of Business & Economic Statistics* 32(2), 206–216.
- Davlasheridze, M., K. Fisher-Vanden, and H. A. Klaiber (2017). The effects of adaptation measures on hurricane induced property losses: Which FEMA investments have the highest returns? *Journal of Environmental Economics and Management* 81, 93–114.
- Dehring, C. A. and M. Halek (2013). Coastal Building Codes and Hurricane Damage. *Land Economics* 89(4), 597–613.
- DeMaria, M., J. Knaff, R. Knabb, C. Lauer, C. Sampson, and R. DeMaria (2009). A new method for estimating tropical cyclone wind speed probabilities. *Weather and Forecasting* 24(6), 1573–1591.
- Deryugina, T. (2017). The Fiscal Cost of Hurricanes: Disaster Aid Versus Social Insurance. *American Economic Journal: Economic Policy* 9(3), 168–198.
- Deryugina, T., L. Kawano, and S. Levitt (2018). The economic impact of hurricane katrina on its victims: evidence from individual tax returns. *American Economic Journal: Applied Economics* 10(2), 202–233.
- Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics* 70, 915–925.
- Doornik, J. A. (2009). Autometrics. In N. Shephard and J. L. Castle (Eds.), *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, pp. 88–121. Oxford University Press.
- Doornik, J. A. and H. Hansen (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 70(s1), 927–939.
- Doornik, J. A. and D. F. Hendry (2013). *Empirical Econometric Modelling*, Volume I of *PcGive 14*. Timberlake Consultants Ltd.
- Emanuel, K. (2005). Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* 436(7051), 686–688.
- Emanuel, K. (2017a). Assessing the present and future probability of hurricane harvey’s rainfall. *Proceedings of the National Academy of Sciences*, 201716222.
- Emanuel, K. (2017b). Will global warming make hurricane forecasting more difficult? *Bulletin of the American Meteorological Society* 98(3), 495–501.

- Engle, R. F. and D. F. Hendry (1993). Testing superexogeneity and invariance in regression models. *Journal of Econometrics* 56(1-2), 119–139.
- Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica* 51(2), 277–304.
- Ericsson, N. R. (2001). Forecast uncertainty in economic modeling. In D. F. Hendry and N. R. Ericsson (Eds.), *Understanding Economic Forecasts*, Chapter 5, pp. 68–92. MIT Press.
- Estrada, F., W. W. Botzen, and R. S. Tol (2015). Economic losses from US hurricanes consistent with an influence from climate change. *Nature Geoscience*.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1), 1.
- Gallagher, J. and D. Hartley (2017). Household finance after a natural disaster: The case of hurricane katrina. *American Economic Journal: Economic Policy* 9(3), 199–228.
- Geiger, T., K. Frieler, and A. Levermann (2016). High-income does not protect against hurricane losses. *Environmental Research Letters* 11(8), 084012.
- Granger, C. W. and Y. Jeon (2004). Thick modeling. *Economic Modelling* 21(2), 323–343.
- Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press.
- Hendry, D. F. and J. A. Doornik (2014). *Empirical model discovery and theory evaluation: automatic selection methods in econometrics*. MIT Press.
- Hendry, D. F. and S. Johansen (2015). Model discovery and Trygve Haavelmo’s legacy. *Econometric Theory* 31(01), 93–114.
- Hendry, D. F., S. Johansen, and C. Santos (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23(2), 337–339.
- Hendry, D. F. and G. E. Mizon (2014). Unpredictability in economic analysis, econometric modeling and forecasting. *Journal of Econometrics* 182(1), 186–195.
- Hendry, D. F. and C. Santos (2010). An Automatic Test of Super Exogeneity. In M. W. Watson, T. Bollerslev, and J. Russell (Eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford University Press.
- Hsiang, S. M. and D. Narita (2012). Adaptation to cyclone risk: Evidence from the global cross-section. *Climate Change Economics* 3(02), 1250011.
- Jarrell, J. D., P. J. Hebert, and M. Mayfield (1992). *Hurricane experience levels of coastal county populations from Texas to Maine*, Volume 46. US Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, National Hurricane Center.

- Johansen, S. and B. Nielsen (2016). Asymptotic theory of outlier detection algorithms for linear time series regression models. *Scandinavian Journal of Statistics* 43(2), 321–348.
- Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *The American Economic Review* 105(3), 1177–1216.
- Kaplan, J., M. DeMaria, and J. A. Knaff (2010). A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Weather and forecasting* 25(1), 220–241.
- 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 on Economic Forecasting*.
- Kellenberg, D. K. and A. M. Mobarak (2008). Does rising income increase or decrease damage risk from natural disasters? *Journal of Urban Economics* 63(3), 788–802.
- Kidder, S. Q., J. A. Knaff, S. J. Kusselson, M. Turk, R. R. Ferraro, and R. J. Kuligowski (2005). The tropical rainfall potential (TRaP) technique. Part i: Description and examples. *Weather and Forecasting* 20(4), 456–464.
- Knutson, T. R., J. L. McBride, J. Chan, K. Emanuel, G. Holland, C. Landsea, I. Held, J. P. Kossin, A. Srivastava, and M. Sugi (2010). Tropical cyclones and climate change. *Nature Geoscience* 3(3), 157–163.
- Kruttli, M., B. Roth Tran, and S. W. Watugala (2019). Pricing poseidon: Extreme weather uncertainty and firm return dynamics. *Available at SSRN 3284517*.
- Letson, D., D. S. Sutter, and J. K. Lazo (2007). Economic value of hurricane forecasts: An overview and research needs. *Natural Hazards Review* 8(3), 78–86.
- Milch, K., K. Broad, B. Orlove, and R. Meyer (2018). Decision science perspectives on hurricane vulnerability: Evidence from the 2010–2012 atlantic hurricane seasons. *Atmosphere* 9(1), 32.
- Mitchell, J. and K. F. Wallis (2011). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics* 26(6), 1023–1040.
- Morana, C. and G. Sbrana (2019). Climate change implications for the catastrophe bonds market: An empirical analysis. *Economic Modelling* 81, 274–294.
- Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31(2), 87–106.
- Murnane, R. J. and J. B. Elsner (2012). Maximum wind speeds and US hurricane losses. *Geophysical Research Letters* 39(16).
- National Science Board (2007). Hurricane Warning: The Critical Need for a National Hurricane Research Initiative. Technical Report NSB-06-115, National Science Foundation.

- NOAA NCEI (2018). U.S. Billion-Dollar Weather and Climate Disasters. <https://www.ncdc.noaa.gov/billions/>. [last accessed 27-September-2018].
- Nordhaus, W. D. (2010). The economics of hurricanes and implications of global warming. *Climate Change Economics* 1(01), 1–20.
- Peters, J., P. Bühlmann, and N. Meinshausen (2016). Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78(5), 947–1012.
- Pielke Jr, R. A., J. Gratz, C. W. Landsea, D. Collins, M. A. Saunders, and R. Musulin (2008). Normalized hurricane damage in the United States: 1900–2005. *Natural Hazards Review* 9(1), 29–42.
- Pielke Jr, R. A. and C. W. Landsea (1998). Normalized hurricane damages in the United States: 1925–95. *Weather and Forecasting* 13(3), 621–631.
- Pretis, F., J. Reade, and G. Sucarrat (2018). Automated General-to-Specific (GETS) Regression Modeling and Indicator Saturation for Outliers and Structural Breaks. *Journal of Statistical Software* 86(1), 1–44.
- Pretis, F., L. Schneider, J. E. Smerdon, and D. F. Hendry (2016). Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator Saturation. *Journal of Economic Surveys* 30(3), 403–429.
- Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 350–371.
- Rappaport, E. N., J. L. Franklin, L. A. Avila, S. R. Baig, J. L. Beven, E. S. Blake, C. A. Burr, J.-G. Jiing, C. A. Juckins, R. D. Knabb, et al. (2009). Advances and challenges at the National Hurricane Center. *Weather and Forecasting* 24(2), 395–419.
- Ravn, M. O. and H. Uhlig (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *The Review of Economics and Statistics* 84(2), 371–376.
- Regnier, E. (2008). Public evacuation decisions and hurricane track uncertainty. *Management Science* 54(1), 16–28.
- Resio, D. T., N. J. Powell, M. A. Cialone, H. S. Das, and J. J. Westerink (2017). Quantifying impacts of forecast uncertainties on predicted storm surges. *Natural Hazards*, 1–27.
- Rossi, B. and T. Sekhposyan (2015). Macroeconomic uncertainty indices based on now-cast and forecast error distributions. *American Economic Review: Papers and Proceedings* 105(5), 650–655.
- Rossi, B., T. Sekhposyan, and M. Soupre (2017). Understanding the sources of macroeconomic uncertainty. Barcelona GSE Working Paper Series, Working Paper No. 920, Barcelona GSE.

- Sadowski, N. C. and D. Sutter (2005). Hurricane fatalities and hurricane damages: are safer hurricanes more damaging? *Southern Economic Journal*, 422–432.
- Shrader, J. (2018). Expectations and adaptation to environmental risks. *mimeo*.
- Smith, A. B. and R. W. Katz (2013). US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Natural Hazards* 67(2), 387–410.
- Strobl, E. (2011). The economic growth impact of hurricanes: evidence from US coastal counties. *Review of Economics and Statistics* 93(2), 575–589.
- U.S. Army Corps of Engineers (2004). Hurricane Assessment Concerns and Recommendations. Available online at: <https://web.archive.org/web/20090726033748/http://chps.sam.usace.army.mil/USHESdata/Assessments/2004Storms/2004-Recommendations.htm> (last accessed December 22, 2017).
- Van de Geer, S., P. Bühlmann, Y. Ritov, R. Dezeure, et al. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics* 42(3), 1166–1202.
- van den Dool, H., J. Huang, and Y. Fan (2003). Performance and analysis of the constructed analogue method applied to US soil moisture over 1981–2001. *Journal of Geophysical Research: Atmospheres* 108(D16).
- Vincenty, T. (1975). Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review* 23(176), 88–93.
- Weinkle, J., C. Landsea, D. Collins, R. Musulin, R. P. Crompton, P. J. Klotzbach, and R. Pielke (2018). Normalized hurricane damage in the continental United States 1900–2017. *Nature Sustainability* 1(12), 808.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48(4), 817–838.
- Willoughby, H. E., E. Rappaport, and F. Marks (2007). Hurricane forecasting: The state of the art. *Natural Hazards Review* 8(3), 45–49.

# A Additional Tables and Figures

**Table A.1: Estimated Terminal Models for a Gauge of 3%**

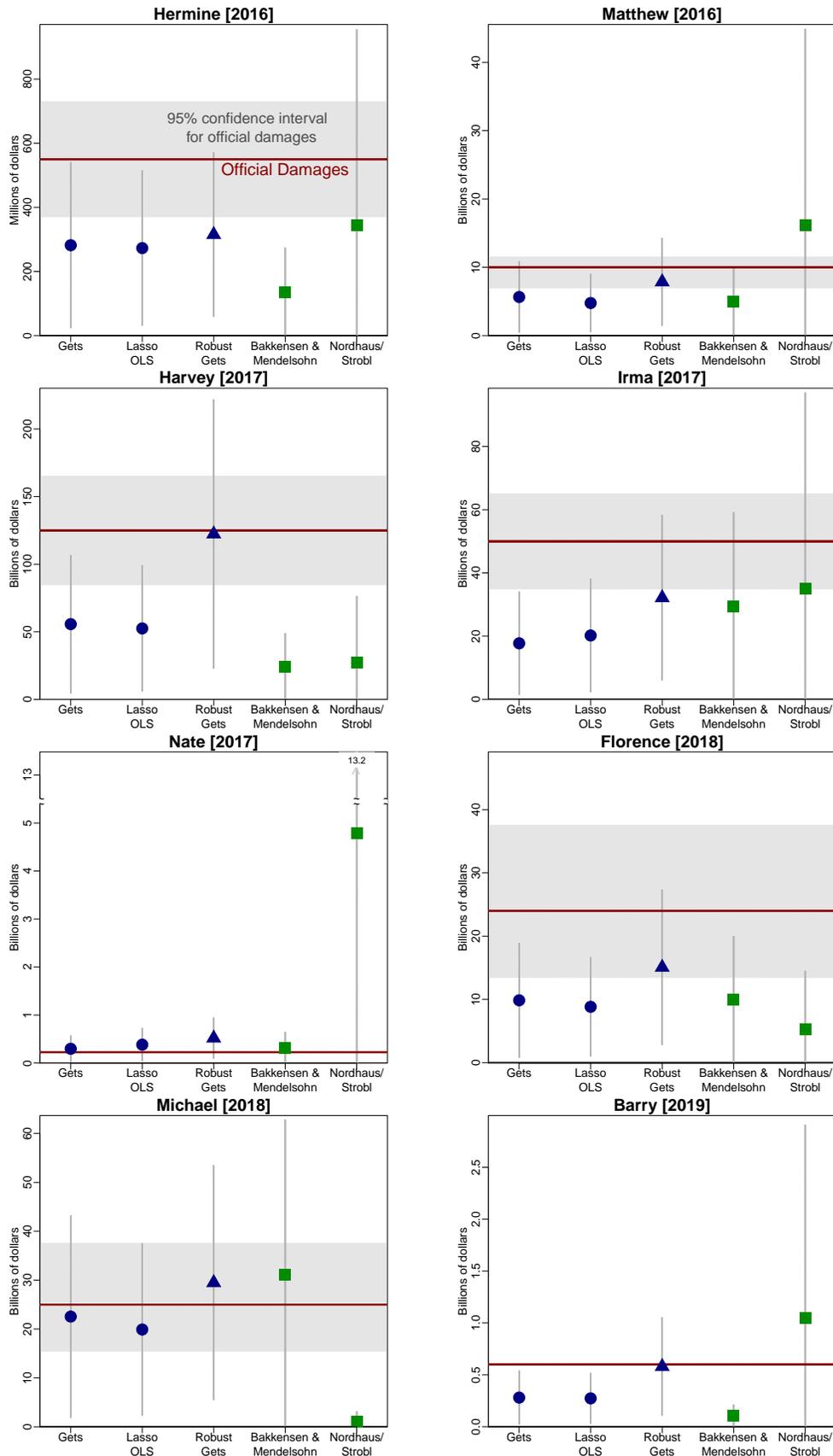
Terminal Model:	Final GUM	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Housing density	0.50** (0.19)	0.58*** (0.14)	0.49*** (0.15)	0.59*** (0.14)	0.56*** (0.14)	0.64*** (0.18)	0.40*** (0.15)	0.70*** (0.18)	0.55*** (0.14)
Income per household	0.85 (0.62)					1.17*** (0.20)	1.30*** (0.20)	1.10*** (0.21)	
Historical fequency	-5.64** (2.55)		-4.85* (2.47)	-5.18** (2.26)		-6.47*** (2.28)		-5.49** (2.44)	-5.05** (2.23)
Max rainfall	0.62* (0.37)	1.12*** (0.32)			1.15*** (0.32)		1.15*** (0.31)		
Max storm surge	1.15*** (0.41)	1.44*** (0.38)		1.45*** (0.39)	1.42*** (0.38)	1.40*** (0.38)	1.34*** (0.38)	1.35*** (0.38)	1.24*** (0.40)
Min central pressure (-)	60.2*** (14.0)	53.9*** (8.66)	80.4*** (13.3)	53.2*** (8.90)	67.2*** (13.5)	50.8*** (8.59)	52.2*** (8.58)	65.5*** (14.1)	54.2*** (8.79)
Max wind speed	-1.02 (1.16)		-0.92 (1.22)		-1.50 (1.12)			-1.32 (1.20)	
Seasonal cyclone energy	0.43* (0.25)			0.60** (0.26)		0.58** (0.25)			0.64** (0.25)
Soil Moisture	0.67 (1.00)		2.27** (0.93)						
Year trend	0.12 (0.10)		-0.07*** (0.01)		-0.08*** (0.01)				-0.07*** (0.01)
Strike trend	0.09 (0.06)	0.05*** (0.01)		0.05*** (0.01)					
AUG	0.48 (0.36)		0.47 (0.39)						
SEP	0.22 (0.33)		0.48 (0.35)						
NY	-0.96 (0.85)					-1.79** (0.74)		-2.23*** (0.77)	
VA	1.66** (0.81)		2.09** (0.86)						2.14** (0.82)
NC	-0.43 (0.45)		-1.39*** (0.38)						-0.80** (0.36)
FL	0.48 (0.34)				0.55* (0.29)				
12-hour forecast error	0.55** (0.24)	0.44* (0.24)	0.25 (0.25)	0.33 (0.24)	0.47* (0.24)	0.46* (0.24)	0.47* (0.24)	0.42* (0.25)	0.39 (0.24)
12-hour radius	2.34 (1.57)	2.35* (1.19)	0.76 (1.19)	2.52** (1.25)	1.62 (1.12)	0.44 (1.02)	0.17 (0.98)	-0.04 (1.01)	1.57 (1.15)
k	17	5	10	6	7	7	5	8	8
Log-likelihood (-)	145.2	157.6	158.4	158.8	156.1	155.0	157.0	157.3	155.1
AIC	3.41	3.42	3.54	3.46	3.43	<b>3.41</b>	3.41	3.48	3.43
HQ	3.65	3.53	3.70	3.58	3.56	3.54	<b>3.52</b>	3.61	3.57
BIC	3.99	3.68	3.93	3.75	3.75	3.73	<b>3.67</b>	3.82	3.77
$\hat{\sigma}$	1.21	1.27	1.32	1.30	1.27	1.26	1.27	1.29	1.26
$R^2$	0.86	0.82	0.82	0.81	0.82	0.83	0.82	0.82	0.83

Notes: Estimated using 98 observations and include a constant and dummy variables for Gerda [1969] and Floyd [1987]. Each terminal model is selected from the Final GUM using a gauge of 3%. Bolded values indicate terminal models with the lowest information criteria. The standard errors are in parentheses.  $k$  is the number of selected regressors. \* $p < 0.1$  \*\* $p < 0.05$  \*\*\* $p < 0.01$

**Table A.2: A Robust Model of Hurricane Damages**

	(1) <i>Gets</i> (3%)	(3) Extension (1%)	(4) Robust
Housing density	0.40*** (0.15)	0.26** (0.10)	0.24** (0.09)
Income per housing unit	1.30*** (0.20)	1.71*** (0.21)	1.76*** (0.19)
Max rainfall	1.15*** (0.31)	0.54** (0.21)	0.56*** (0.20)
Max storm surge	1.34*** (0.38)	0.95*** (0.25)	0.98*** (0.24)
Min central pressure (-)	52.2*** (8.58)	56.4*** (5.61)	57.8*** (5.41)
Forecast errors	0.47* (0.24)	0.26 (0.16)	0.27* (0.15)
Ex-ante Uncertainty	0.17 (0.98)	1.94* (1.06)	2.10** (0.96)
Income per housing unit sq.		0.65*** (0.14)	0.44*** (0.09)
Outlying storms:			-3.19*** (0.35)
Helene [1958]		-2.59*** (0.85)	
Cindy [1959]		-4.09*** (0.85)	
Gracie [1959]		-2.61*** (0.83)	
Alma [1] [1966]		-4.28*** (0.83)	
Bret [1999]		-2.45*** (0.85)	
Alex [2004]		-3.20*** (0.84)	
Arthur [2014]		-3.36*** (0.98)	
$\hat{\sigma}$	1.267	0.804	0.795
$R^2$	0.821	0.934	0.930
$\chi_{nd}^2(2)$	7.93** [0.019]	0.43 [0.806]	0.65 [0.722]
$F_{Het}$	2.00** [0.028]	1.27 [0.237]	0.87 [0.617]
$F_{Het-X}$	1.37 [0.139]	0.92 [0.615]	0.81 [0.769]
$F_{RESET23}$	1.40 [0.253]	1.40 [0.253]	1.51 [0.226]

*Notes:* All equations are estimated using 98 observations and include a constant and dummy variables for Gerda [1969] and Floyd [1987]. The standard errors are in parentheses. The diagnostic tests are the  $\chi_{nd}^2(2)$  test for non-Normality (Doornik and Hansen, 2008), the  $F_{Het} / F_{Het-X}$  test for residual heteroskedasticity (with and without cross products; White, 1980), and the  $F_{RESET23}$  test for incorrect model specification (Ramsey, 1969). The tail probability associated with the null hypothesis of each diagnostic test statistic is in square brackets. Income per housing unit is demeaned to facilitate interpretability of the coefficients. \* $p < 0.1$  \*\* $p < 0.05$  \*\*\* $p < 0.01$



Notes: Official damages and confidence intervals are from NOAA's Tropical Cyclone Reports and NOAA's Billion-Dollar Events Database. The forecast standard errors are computed using the delta method under the assumption of normality.

Figure A.1: Out-of-sample damage 'predictions' by storm and method

## B Descriptions of Other Variables

County level population and personal income estimates come from the Bureau of Economic Analysis (BEA) since 1969. Prior to 1969, I use county level population estimates from the U.S. Census, available for each decade, along with state level population and personal income estimates available annually from the BEA. I compute annual county level population values prior to 1969 by interpolating county level population shares between decades and then distributing them using state level data. I use a similar approach for computing land area and housing units associated with each strike.<sup>10</sup>

Annual county level personal income prior to 1969 is estimated as follows. First, I assume that county level personal income shares were constant from 1955 to 1969. Second, I estimate annual income shares using a fixed effects panel data model and then starting in 1969, backcast the shares to 1955.<sup>11</sup> I combine the shares with state level income to get a county level estimate. I average these two approaches to get a robust measure.

I compute the real-time measure of historical strike frequency using county level hurricane strikes since 1900. The strike frequency for a county in a given year is computed over time by taking the number of hurricanes that struck that county since 1900 divided by the number of years that have passed. Strike level historical frequency is computed as an average of the strike frequencies of all counties struck by the hurricane at the time of the strike.

Since damages are measured at the strike level, I aggregate all variables across impacted counties, which alleviates concerns about county level estimates but requires me to choose which counties are impacted. I focus on coastal counties (Jarrell et al., 1992), which overweights the importance of the coast but is less likely to over-weight the impact of wind damages as in Strobl (2011), Hsiang and Narita (2012), and Deryugina (2017).

The maximum sustained (1-minute) surface (10 meter) wind speed, minimum central pressure at landfall, maximum storm surge height, and accumulated seasonal cyclone energy are all obtained for each hurricane from NHC tropical cyclone reports. The maximum rainfall for each strike comes from NOAA's Weather Prediction Center.

<sup>10</sup>Pielke Jr et al. (2008) use a similar approach. I aggregate counties using BEA's modifications to Census codes: <https://www.bea.gov/regional/pdf/FIPSMODIFICATIONS.pdf> (last accessed November, 2016).

<sup>11</sup>The panel data model was estimated over for all U.S. counties from 1969 to 1999 using leads of income shares and population shares as explanatory variables.

**Table B.1: Variable Descriptions and Summary Statistics**

Variable	Description	Years	Min	Average	Max	Source
<b>Damages (D)</b>						
DAMAGE	Nominal Damage (U.S. \$1,000 )	1955-2015	\$28	\$3,990,506	\$105,900,000	NOAA, NHC
<b>Socio-economic Vulnerabilities (V)</b>						
IP	Income Per Capita (\$ per person)	1955-2015	\$864	\$16,430	\$60,213	BEA
HD	Housing Unit Density (houses per acre)	1955-2015	5	104	1,672	Census
IH	Income Per Housing Unit (\$1,000 per unit)	1955-2015	\$2	\$37	\$140	BEA, Census
FREQ	Historical Hurricane Frequency (Average per year)	1955-2015	0.01	0.09	0.32	NOAA, HRD
LEVEE	Levee Length Density (miles per acre)	1955-2015	0	0.03	0.22	USACE, NLD
<b>Natural Forces (F)</b>						
WIND	Max Sustained Wind Speed (kt)	1955-2015	65	90.3	150	NOAA, HRD
PRESS	Central Pressure at Landfall (mb)	1955-2015	909	965	1003	NOAA, HRD
RAIN	Max Rainfall (in)	1955-2015	4.8	13.75	38.5	NOAA, WPC
SURGE	Max Surge (ft)	1955-2015	0	8.5	27.8	NOAA, NHC
ACE	Accumulated Cyclone Energy (Seasonal)	1955-2015	17	135	250	NOAA, HRD
MOIST	Deviations from trend soil moisture (in)	1955-2015	-4.75	1	5.7	NOAA, ESRL
GST	Land, Air and Sea-Surface Temp. index	1955-2015	0.1	0.34	0.93	NASA, GISS
<b>Forecast Accuracy and Uncertainty (U)</b>						
FORC12	12-Hour Official Track Error (nautical miles)	1955-2015	5	34	114	NOAA, NHC
RADII12	Implied 12-hour radius of uncertainty (nautical miles)	1955-2015	34	70	59	NOAA, NHC
NAIVE12	12-Hour Naïve Track Error (nautical miles)	1970-2015	5	31	97	NOAA, NHC
SKILL12	Ratio of 12-Hour naïve forecast error to FORC12	1970-2015	0.19	1.44	10.35	NOAA, NHC
WARN	Warning time over coast length (100 hours per mile)	1955-2015	0.7	11.33	45.78	NOAA, NHC

Notes: NOAA: National Oceanic and Atmospheric Administration; NHC: National Hurricane Center; HRD: Hurricane Research Division; WPC: Weather Prediction Center; ESRL: Earth System Research Laboratory; NASA: National Aeronautics and Space Administration; GISS: Goddard Institute of Space Studies; BEA: Bureau of Economic Analysis; Census: U.S. Census Bureau; USACE, NLD: US Army Corps of Engineers, National Levee Database.

Model-based estimates of monthly soil moisture, derived using methods devised by van den Dool et al. (2003), are obtained from NOAA’s Earth System Research Laboratory. These estimates are then linked in the nearest grid point to a county’s centroid. County estimates are averaged across impact counties for each hurricane strike and then smoothed.<sup>12</sup> I use the smoothed estimate for the month prior to the strike.

I compute storm-level estimates of sea surface air temperature following Estrada et al. (2015) using data from NASA’s global mean surface temperature index based on land-surface air temperature anomalies. The monthly series is smoothed using the Hodrick-Prescott filter with  $\lambda = 129,600$ . The estimate for the month prior to the hurricane strike is used. Table B.1 and provide sources and summary statistics for each of the variables.

<sup>12</sup>I use the Hodrick-Prescott filter and set the smoothing parameter equal to 129,600 following Ravn and Uhlig (2002) for monthly data.

## C Technical Appendix

Consider a simple representation of damages as:

$$\ln(p_i y_i) = X_i' \beta + \epsilon_i, \quad (\text{C.1})$$

where  $X_i$  is a vector of explanatory variables,  $\epsilon_i$  is assumed to be iid normal and  $\mathbb{E}(X_i \epsilon_j) = \mathbf{0} \forall i, j$ . I estimate this model, apply the delta method, and re-scale by prices to get a prediction of real damages

$$\widehat{y}_i \sim \text{IN} \left( y_i, (y_i \sigma_{i,\widehat{y}})^2 \right). \quad (\text{C.2})$$

The difference between this prediction and some counterfactual,  $\widetilde{y}_i = \kappa_i \widehat{y}_i$  for  $\kappa_i \neq 0$ , gives

$$(\widetilde{y}_i - \widehat{y}_i) \sim \text{IN} \left( (\kappa_i - 1) y_i, \mathbb{V}(\widetilde{y}_i - \widehat{y}_i) \right), \quad (\text{C.3})$$

$$\mathbb{V}(\widetilde{y}_i - \widehat{y}_i) = (\kappa_i y_i \sigma_{i,\widehat{y}})^2 + (y_i \sigma_{i,\widehat{y}})^2 - \text{Cov}(\widetilde{y}_i, \widehat{y}_i) - \text{Cov}(\widehat{y}_i, \widetilde{y}_i),$$

From the independence, normality and exogeneity assumptions, (C.3) can be cumulated as:

$$\sum_{i=1}^n (\widetilde{y}_i - \widehat{y}_i) \sim \text{N} \left( \sum_{i=1}^n (\kappa_i - 1) y_i, \sum_{i=1}^n \mathbb{V}(\widetilde{y}_i - \widehat{y}_i) \right). \quad (\text{C.4})$$

The variance is the residual error variance plus parameter estimation uncertainty which Doornik and Hendry (2013) show can be estimated as

$$\widehat{\sigma}_{i,\widehat{y}}^2 = X_i' \left[ \widehat{\mathbb{V}}(\widehat{\beta}) \right] X_i + \widehat{\sigma}_{i,\epsilon}^2 \quad (\text{C.5})$$

$$\widehat{\sigma}_{i,\widetilde{y}}^2 = \widetilde{X}_i' \left[ \widehat{\mathbb{V}}(\widehat{\beta}) \right] \widetilde{X}_i + \widehat{\sigma}_{i,\epsilon}^2, \quad (\text{C.6})$$

where  $\mathbb{V}(\widehat{\beta})$  is the covariance matrix of parameter estimates. The covariance is estimated as

$$\widehat{\text{Cov}}(\widetilde{y}_i, \widehat{y}_i) = \kappa_i y_i^2 \left\{ \widetilde{X}_i' \left[ \widehat{\mathbb{V}}(\widehat{\beta}) \right] X_i + \widehat{\sigma}_{i,\epsilon}^2 \right\}. \quad (\text{C.7})$$

Bringing these pieces together and simplifying, the total estimate of the variance is

$$\widehat{\mathbb{V}}(\widetilde{y}_i - \widehat{y}_i) = y_i^2 \left\{ (\kappa_i \widetilde{X}_i - X_i)' \left[ \widehat{\mathbb{V}}(\widehat{\beta}) \right] (\kappa_i \widetilde{X}_i - X_i) + (\kappa_i - 1)^2 \widehat{\sigma}_{i,\epsilon}^2 \right\}. \quad (\text{C.8})$$

The variance is driven by the ratio between damage predictions ( $\kappa_i$ ) and the difference between the regressors ( $\widetilde{X}_i - X_i$ ) and is exacerbated by damages, the residual error variance and the covariance of parameter estimates. By construction when  $\kappa_i = 1$  then  $\widetilde{X}_i = X_i$  so the total variance collapses to zero.