Andrew B. Martinez
Nationality: United States of America
https://sites.google.com/view/andrewbmartinez

Magdalen College
Oxford, United Kingdom
andrew.martinez@economics.ox.ac.uk

UNIVERSITY OF OXFORD

Placement Officer:	Johannes Abeler	johannes.abeler@economics.ox.ac.uk	+44 1865 281440
Graduate Administrator:	Julie Minns	julie.minns@economics.ox.ac.uk	+44 1865 281162

EDUCATION

2015– *DPhil. (PhD); Economics (expected 2019)*
UNIVERSITY OF OXFORD, MAGDALEN COLLEGE

2013–15 *MPhil.; Economics, July 2015*
UNIVERSITY OF OXFORD, ST. CROSS COLLEGE

2008–10 *M.A.; International Trade & Investment Policy, May 2010*
THE GEORGE WASHINGTON UNIVERSITY, ELLIOTT SCHOOL OF INTERNATIONAL AFFAIRS

2003–07 *B.A., cum laude; Economics, International Studies, German Language & Literature, May 2007*
GUILFORD COLLEGE

REFERENCES

David F. Hendry
Nuffield College, University of Oxford
david.hendry@nuffield.ox.ac.uk

Jennifer L. Castle
Magdalen College, University of Oxford
jennifer.castle@magd.ox.ac.uk

Neil R. Ericsson
Federal Reserve Board
Department of Economics, GWU
neil.r.ericsson@frb.gov

Bent Nielsen
Nuffield College, University of Oxford
Department of Economics, University of Oxford
bent.nielsen@nuffield.ox.ac.uk

RESEARCH AND TEACHING FIELDS

Econometrics, Time Series Econometrics, Applied Macroeconomics, International Economics

TEACHING EXPERIENCE

Teaching Assistant, Department of Economics, University of Oxford, UK (2015-present)
MPhil (graduate) Econometrics I: OLS, IV, GMM, MLE (Fall: 2015-2018)
MPhil (graduate) Econometrics II: Macroeconometrics (Spring: 2016-2018)
MPhil (graduate) Econometrics III: Microeconometrics (Spring: 2018)

Instructor, OxMetrics Training Course, Motability Operations, UK (Sep. 2018)

Teaching Assistant, Economic Forecasting Summer School, ISF Boulder, USA (June 2018)

Instructor, Econometric Modeling and Forecasting, Universidad de La Habana, CU (Apr. 2016)

Teaching Assistant, Econometrics Spring School, George Washington University, USA (Mar. 2016)

Teaching Assistant, Econometrics Summer School, Aix-Marseille University, FR (Sep. 2015)

Tutor, Macroeconomics & International Economics, Guilford College, USA (2005-2007)

Research Experience and Other Employment

2018	Summer Associate, Global Investment Research Division, Goldman Sachs International
2017	Dissertation Intern, Research Department, Federal Reserve Bank of Cleveland
2017	Short Term Expert, Institute for Capacity Development, International Monetary Fund
2016	Research Assistant, Professor Sophocles Mavroeidis, University of Oxford
2014–15	Research Assistant, Professor Sir David Hendry, University of Oxford
2010–13	Research Officer, Independent Evaluation Office, International Monetary Fund
2009	Research Assistant, Professor Graciela Kaminsky, The George Washington University
2008–10	International Economist, Office of Economics, U.S. International Trade Commission
2008	Enterprise Researcher, Research Department, RainKing Software Inc.
2007	Summer Research Associate, Maine International Trade Center
2006	Political and Economic Affairs Intern, U.S. Consulate Leipzig, U.S. Department of State

Honors, Scholarships, and Fellowships

2018-19	University of Oxford, David Walton Distinguished Doctoral Scholarship
2018	First Runner-up, Best Student Presentation, International Symposium on Forecasting
2017	AEA Summer Economics Fellow
2016-19	Oxford Martin School Fellow
2015-18	Robertson Foundation Doctoral Scholarship
2015–18	University of Oxford, Department of Economics Doctoral Scholarship
2015	International Institute of Forecasters Student Forecasting Award

Publications

1. "Evaluating Forecasts, Narratives and Policy using a Test of Invariance" (with Jennifer L. Castle and David F. Hendry), *Econometrics*, Vol. 5, 3 (2017): 39, dx.doi.org/10.3390/econometrics5030039

2. "Evaluating Multi-Step System Forecasts with Relatively Few Forecast-Error Observations" (with David F. Hendry), *International Journal of Forecasting*, Vol. 33, 2 (2017), pp. 359-372, dx.doi.org/10.1016/j.ijforecast.2016.08.007

3. "How Good Are U.S. Government Forecasts of the Federal Debt?" *International Journal of Forecasting*, Vol. 31, 2 (2015), pp. 312-324, dx.doi.org/10.1016/j.ijforecast.2014.08.014

4. "Overview of U.S.-China Trade in Advanced Technology Products" (with Alexander Hammer and Robert Koopman), *Journal of International Commerce and Economics*, Vol. 3 (2011), pp. 1-16

Job Market Paper and Research in Progress

"A False Sense of Security: The Impact of Forecast Accuracy on Hurricane Damages" (JMP)

Can forecasts of natural disasters alter their destructiveness? Poor forecasts increase damages when individuals do not mitigate risks based on the false belief that they will be unaffected. We test this hypothesis by examining the impact of 12-hour-ahead forecasts on hurricane damages and find that larger errors in the storm's predicted landfall location lead to higher damages. The cumulative reduction in damages from forecast improvements since 1970 is about $82 billion. This exceeds the U.S. government's spending on these forecasts and private willingness to pay for them. The benefits from forecast improvements are underestimated and individual adaptation decisions matter.

"Testing for Differences in Path Forecast Accuracy: The Dynamics Matter"

The trajectory and path of future outcomes play a crucial role in policy decisions. However, analyses of forecast accuracy typically focus on point forecasts. The path forecasts provide additional insight into the forecast dynamics. I propose a test for differences in path forecast accuracy using the link between path forecast evaluation metrics and the joint predictive density. This path test nests and extends existing joint multi-horizon forecast accuracy tests. Simulations highlight the trade-offs when using path forecast accuracy tests to detect a broad range of differences in the forecasts. I compare the Federal Reserve's Greenbook path forecasts against four DSGE model forecasts. The results show that differences in forecast dynamics play an important role in the assessment of path forecast accuracy.

Permanent Working Papers and Reports

1. "On the Accuracy and Efficiency of IMF Forecasts: A Survey and some Extensions" (with Hans Genberg), *IEO Background Paper,* BP/14/04, February 2014

2. "The IMF-WEO Forecast Process" (with Hans Genberg and Michael Salemi), *IEO Background Paper,* BP/14/03, February 2014

3. "IMF Bilateral Surveillance on International Reserves" (with Angana Banerji), *IEO Background Paper,* BP/12/02, August 2012

4. "Comparing Government Forecasts of the United States' Gross Federal Debt," *GWU Research Program on Forecasting Working Paper*, February 2011

5. "CoRe NTMs Database: A Compilation of Reported Non-Tariff Measures" (with Jesse Mora and Jose Signoret), *USITC Office of Economics Working Paper*, December 2009

Languages: *Human:* English (native), German (fluent); *Programming*: EViews, Ox, R, Stata

Member: American Economic Association (2018-present), The Econometric Society (2018-present), International Institute of Forecasters (2012-present), Royal Economic Society (2017-present)

Presentations: *2018*: 2nd Conference on Forecasting at Central Banks, 20th OxMetrics Users Conference, 3rd Conference on Econometric Models of Climate Change, 38th International Symposium on Forecasting, European Geophysical Union Annual Meetings, VU Amsterdam Econometrics Lunch Seminar, VU Amsterdam Institute for Environmental Studies Lunch Seminar; *2017*: RES PhD Meetings, Computational and Financial Econometrics, Oxford Econometrics Lunch Seminar, Federal Reserve Bank of Cleveland Brown Bag Seminar, 37th International Symposium on Forecasting, 6th International Summit on Hurricanes and Climate Change, Oxford Econometrics Lunch Seminar; *2016*: INET Oxford Research Seminar, Econometric Models of Climate Change, 18th OxMetrics Users Conference, 36th International Symposium on Forecasting, 17th OxMetrics Users Conference; *2015*: INET Oxford Research Seminar, 16th OxMetrics Users Conference; *2012*: 32nd International Symposium on Forecasting, 11th OxMetrics Users Conference; *2011*: Southern Economic Association Annual Meetings, Federal Forecasters Conference, GWU Brown Bag Seminar on Forecasting; *2007*: N.C. Undergraduate History Thesis Writers Conference

Referee: *Empirical Economics*, *International Journal of Forecasting*, *International Review of Economics and Finance*, *Journal of Applied Econometrics*, *Journal of Econometrics*

A False Sense of Security: The Impact of Forecast Accuracy on Hurricane Damages

Andrew B. Martinez*

March 26, 2019

link to latest version

Abstract

Can forecasts of natural disasters alter their destructiveness? Poor forecasts increase damages when individuals do not mitigate risks based on the false belief that they will be unaffected. We test this hypothesis by examining the impact of 12-hour-ahead forecasts on hurricane damages and find that larger errors in the storm's predicted landfall location lead to higher damages. The cumulative reduction in damages from forecast improvements since 1970 is about $82 billion. This exceeds the U.S. government's spending on these forecasts and private willingness to pay for them. The benefits from forecast improvements are underestimated and individual adaptation decisions matter.

Keywords: Adaptation, Model Selection, Natural Disasters, Uncertainty

JEL classifications: C51, C52, Q51, Q54

*Department of Economics and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, andrew.martinez@economics.ox.ac.uk. This research was supported in part by a grant from the Robertson Foundation (grant 9907422). Thanks to Laura Bakkensen, Maria Balgova, Nadia Bloemendaal, Jennifer L. Castle, Todd E. Clark, Tatyana Deryugina, Jurgen Doornik, Neil R. Ericsson, David F. Hendry, Luke P. Jackson, Felix Pretis and to 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 for helpful comments and suggestions. Special thanks to Robert G. Williams for sparking my interest in hurricanes and their forecasts, to Luke P. Jackson for help with processing the soil moisture data and to Alexandra Woodruff for help with obtaining the CRS data. Numerical results were obtained using OxMetrics 7.2 (OSX/U), PcGive 14.2 and R 3.4.3. Additional data processing was performed using Stata 15.1. The figures were generated using OxMetrics 7.2 and R 3.4.3. The manuscript was prepared with LyX 2.2.3. All errors are my own.

1 Introduction

Damages from natural disasters in the United States, driven in part by several hurricanes, 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 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). Despite their widespread effects, natural disasters are localized events. Their costs are determined in part by individuals' decisions about how and when to protect their property. Decisions about whether to board up windows, stack sandbags, harvest crops, relocate property etc. have to be made in advance. They rely on forecasts of the event's occurrence, location, and severity produced up to several days ahead. These forecasts, despite dramatic improvements, are far from perfect and can exhibit large errors, which are often not realized until it is too late. Large and unexpected forecast errors, even up to just a few hours ahead, may lead people in the disaster area to protect their property less than they would have otherwise. Individuals who place too much faith in the accuracy of the forecasts may make poor decisions which result in higher damages.

This paper seeks to quantify, how the accuracy of short-term forecasts leading up to a natural disaster affect the resulting destruction. While a considerable amount of research examines changes in the natural hazards and vulnerabilities associated with natural disasters, little attention has been paid to the effectiveness of short-term damage mitigation decisions just prior to a disaster's occurrence. Even less research focuses on how early warning systems based on short-term forecasts of the event affects these decisions. Quantifying the extent to which short-term forecasts of natural disasters help individuals make better damage mitigation decisions is important for understanding the effectiveness of short-term adaptation efforts, for illuminating the drivers of the rising costs of natural disasters, and for understanding the speed of the ultimate economic recovery.

We start by showing that, in an expected utility framework, errors in the forecasts of a natural disaster can affect damages if individuals place too much faith in the accuracy of these forecasts when formulating their beliefs about the benefits of damage mitigation.

Next, using our theoretical framework as a guide we test for evidence of these effects using a newly constructed database of all hurricanes to strike the continental United States from 1955-2014. We formulate an empirical model of damages with many determinants including the 12-hour-ahead landfall-forecast errors and estimate it using ordinary least squares. Then, we use model selection methods to determine the best model specification and also whether forecast errors are among the most important determinants of damages. The robustness of our results is confirmed using a variety of selection procedures and model specifications. Finally, we conduct a counterfactual exercise in order to quantify how much improvements in short-term forecast accuracy since 1970 have altered hurricane damages.

Short-term forecast errors of the storm's location, together with a handful of other variables, explain most of the variation in aggregate hurricane damages over the past sixty years. A one standard deviation increase in the distance between where a hurricane is expected to strike and where it actually strikes is associated with up to $\$3,000$ in additional damages per household affected by the hurricane. Interpreting this through the lens of our theoretical framework indicates that short-term forecasts guide individual damage mitigation decisions. In aggregate, individual decisions to protect and relocate property in the face of a disaster have a significant impact on the overall costs.

The U.S. government devotes considerable resources to improving its hurricane forecasts despite limited evidence of their value beyond reductions in fatalities. We quantify the short-term reduction in damages due to hurricane forecast improvements. The predicted cumulative damages prevented due to improvements in forecast accuracy since 1970 is about \$82 billion. This means that the cumulative net benefit is between $\$30 - 71$ billion after accounting for what the U.S. federal government spends on hurricane operations and research. The benefits of further forecast improvements also outweigh measures of society's 'willingness to pay' for them (see Katz and Lazo, 2011).

This paper contributes to several strands of the literature. The first is related to the measurement of forecast uncertainty. Analyses of uncertainty go back at least as far as Knight (1921), who distinguished between risk which is measurable ex-ante and 'Knightian' uncertainty which is not. Recent work defines uncertainty predominantly in terms of second moments; see Bloom (2009, 2014). This is less true for forecast uncertainty, since as Jurado et al. (2015, p. 1178) argue, "what matters for economic decision making is [...] whether the economy has become more or less predictable; that is, less or more uncertain." We extend Rossi and Sekhposyan (2015)'s measure of forecast error uncertainty to allow for time-varying densities. We show how and under what conditions we can approximate this measure as the ratio of forecast errors to their ex-ante standard deviation. This allows us 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.

We also contribute to the literature on the effectiveness of adaptation to natural disasters. Our work is related to Bakkensen and Mendelsohn (2016), who estimate a model of global tropical cyclone damages to understand the relationship between income and adaptation in natural disasters. Although they find evidence of adaptation globally, they argue that the United States is an exception. We extend their approach by allowing forecast errors to alter adaptation decisions. This allows us to measure short-term adaptation efforts separately from income. We find evidence of adaptation in the United States while also confirming the finding that higher income is not associated with lower damages.

Finally, we also contribute to the literature on the value of environmental forecasts. Our work relates to a large number of studies including Krzysztofowicz and Davis (1983), Carsell et al. (2004), Regnier (2008) and Pappenberger et al. (2015). When assessing the value of hurricane forecasts, previous studies typically focus on the value of improved evacuation decisions and reduced fatalities. Our analysis illustrates that the benefits of accurate forecasts also play an important role in short-term damage mitigation decisions. This allows us to consider the cost of forecast improvements and the private willingness to pay for those improvements in a different context and we show that the benefits to forecast improvements are even higher than previously considered.

The rest of the paper is structured as follows: Section 2 explores the link between natural disasters, damages, and uncertainty. Section 3 proposes a theoretical model of how forecast uncertainty can impact damages. The rest of the paper applies this model to an application of damages from hurricane strikes where section 4 describes the statistical methods and the data used. Section 5 presents the results while section 6 assesses their robustness. Section 7 discusses the implications of improving the forecasts and section 8 concludes.

2 Natural Disasters: Damages and Forecasts

The destruction from a wide range of natural disasters, including wild fires, hurricanes, tornadoes, droughts, and floods, reached a record breaking $313 billion in the United States in 2017 (NOAA NCEI, 2018). This likely underestimates the total cost as natural disasters have both persistent and transitory economic impacts. Strobl (2011) argues that a one percent increase in the direct cost of a natural disaster is associated with a transitory decline in local economic growth by 0.45 percentage points. Baker and Bloom (2013) also find that natural disasters precede declines in economic activity. Deryugina (2017) finds that natural disasters are associated with additional widespread direct and indirect fiscal costs. Gallagher and Hartley (2017) find that Hurricane Katrina [2005] altered borrowing patterns and spurred efforts to deleverage, while Deryugina et al. (2018) finds that it led to lasting changes on where people live but only temporary changes otherwise. This illustrates that natural disasters can have widespread and lasting economic impacts.

There are many potential determinants of the destructiveness of a natural disaster. Natural hazards such as the maximum wind speed of a hurricane, its storm surge, rainfall, and minimum central pressure are considered to be the most important; see Nordhaus (2010), Murnane and Elsner (2012) and Chavas et al. (2017). Damages are also determined by the vulnerabilities of a location. This is often measured by how much income, housing or capital stock there is in an area; see Pielke Jr and Landsea (1998), Pielke Jr et al. (2008), and Neumayer and Barthel (2011). There is also a growing literature on medium and longer-term efforts to mitigate damages. This is captured through higher incomes, building codes, and spending on government damage mitigation programs; see Bakkensen and Mendelsohn (2016), Geiger et al. (2016), Dehring and Halek (2013) and Davlasheridze et al. (2017).

Little attention has been paid to the role that early warning systems for natural disasters and their forecasts could play in altering the destructiveness. For example, Deryugina et al. (2018, p. 202) claims that Hurricane Katrina [2005] "struck with essentially no warning", which ignores that there were warnings in place several days before the storm struck. On the other hand, Letson et al. (2007) argue that there is a trade-off between damage mitigation efforts and forecast improvements in the medium-term. In their view, earlier warnings and better forecasts of a natural disaster may lead to moral hazard concerns as people are more likely to move into higher risk locations due to the declining risk of fatalities. This in turn would lead to increased vulnerabilities and higher damages. Sadowski and Sutter (2005) claim that the negative bi-variate correlation between damages and fatalities and between damages and forecast errors (see Appendix Table A.1) provides evidence in support of this argument. However, these simple relationships are most likely confounded by the longer-term trends in technological change and economic growth.

We focus instead on the short-term impact of forecasts on damages. We argue that forecasts matter because they are used for planning and mitigation decisions. Behavioral response surveys conducted by the U.S. Army Corps of Engineers (2004) illustrate this link:

> "Many people believe the storm will miss their location, sometimes placing too much faith in the forecast track of the storm, and sometimes those misconceptions are reinforced by similar misconceptions by emergency management officials. In some cases, 40% of the respondents said they have never spent anything to make their homes safer in hurricanes".

This finding, which is reaffirmed in more recent surveys (Milch et al., 2018), suggests that individuals do not mitigate potential damages in part due to the forecasts. Individuals form their beliefs and make damage mitigation decisions based on imperfect information. This provides a basis though which poor forecasts can increase damages. The next section formalizes this relationship within a theoretical model of damages from natural disasters.

3 Theoretical Framework

We start by developing a theoretical framework to illustrate how damage mitigation decisions can provide a link between forecasts and damages. Consider the expected expenditures when facing the risk of a natural disaster. Damages, $d(\cdot)$, occur with probability $p(\cdot)$, and a cost $c(\cdot)$ of mitigating them. Damages depend on the event's location and severity, F, the vulnerable assets, V, and any mitigation efforts undertaken, A.

We assume that the sequence of actions are: (a) nature determines the location and severity of the event, F; (b) individuals form imperfect beliefs about the event, $\hat{F}$, and the probability $p(\hat{F})$ of damages $d(\cdot)$; (c) they then choose a level of mitigation, A, based on their beliefs, the likely damages, and costs; (d) the event occurs based on nature's predetermined location and severity. Although individuals are not observed directly, we assume that they face similar choices based on publicly available information about the event and so the aggregate decisions across all individuals can be modeled as a single representative agent.

We can write the choice in (c) as an expenditure minimization problem under uncertainty:

$$\min_A p(\hat{F})d(V, \hat{F}, A) + c(A). \tag{3.1}$$

This formulation dates back to Von Neumann and Morgenstern (1945) where the probability of the event is known. Modifications adapt this formulation to allow for the probability to be unknown; see Kahneman and Tversky (1979) among others. The typical setup allows for individuals to be ex-ante uncertain (or ambiguous) about probabilities by integrating over all possible future states. A false sense of security occurs when there is no ex-ante uncertainty and where probabilities are determined exclusively by point forecasts of the event.

We denote the expenditure function in (3.1) by $E(A)$. Taking the derivative w.r.t A gives

$$\frac{\partial E(A)}{\partial A} = E'(A) = p(\hat{F})d'(V, \hat{F}, A) + c'(A). \tag{3.2}$$

Assuming that $d(\cdot)$ is convex in A is necessary and sufficient for a solution. Then the optimal level of mitigation A^* (conditional on ex-ante beliefs about $\hat{F}$) satisfies

$$d'(V, \hat{F}, A^*) = -\frac{c'(A^*)}{p(\hat{F})}, \tag{3.3}$$

where we assume that $0 < c'(A) < \infty$. This implies that individuals should mitigate up until the marginal benefit of mitigation is equal to the ratio between the marginal cost of mitigation and the perceived probability of damages occurring.[1] Thus, mitigation choice is a function of individuals' beliefs about the event which we denote as: $A(\hat{F})$.

Even when beliefs are formed without any ex-ante uncertainty, they can be wrong or uncertain ex-post. We can assess the impact of this ex-post uncertainty by examining the difference between the ex-ante and ex-post optimal marginal benefit of mitigation. If mitigation is optimal in both cases, then with some algebraic manipulation this is written as:

$$d'(V, F, A(F)) - d'(V, \hat{F}, A(\hat{F})) = \frac{1}{p(\hat{F})}\left[c'(A(\hat{F})) - c'(A(F))\right] + \left[\frac{p(F) - p(\hat{F})}{P(\hat{F})}\right]\frac{c'(A(F))}{p(F)}. \tag{3.4}$$

The first term in (3.4) represents the difference in the ex-post and ex-ante marginal costs and the second term is driven by the error in the perceived probability of damages.

[1] This is analogous to the classic optimal demand for insurance problem.

From (3.4), the marginal reduction in damages is lower ex-post than it is ex-ante if and only if $\frac{c'(A(\hat{F}))-c'(A(F))}{c'(A(F))} > \frac{p(\hat{F})-p(F)}{p(F)}$. When marginal costs are constant or increasing then $p(\hat{F}) < p(F)$ is required. If this holds, then the marginal benefit of additional mitigation is greater ex-ante than it is ex-post. Thus, when $\hat{F} < F$, mitigation will be lower, $A(\hat{F}) < A(F)$, and damages higher, $d(V, F, A(F)) \leq d(V, F, A(\hat{F}))$, than they otherwise would have been.

If mitigation choices are sub-optimal, then the marginal cost is not a binding constraint and (3.4) does not hold. However, a second order Taylor series expansion of $d(V, F, A(\hat{F}))$ around F illustrates that as long as damages are convex in adaptation then $d(V, F, A(F)) < d(V, F, A(\hat{F}))$ if $\hat{F} < F$. Thus, irrespective of the optimality of mitigation, our theoretical framework predicts that higher ex-post uncertainty about an event in the context of an ex-ante optimistic bias is associated with lower levels of adaptation and higher damages.

This framework provides a guide for thinking about a realistic model of damages from natural disasters. We assume, as is common in the literature, that damages can be represented by a Cobb-Douglas power function. Taking logarithms gives the expression:

$$ln(d_i) = c + \alpha ln(\boldsymbol{V}_i) + \beta ln(\boldsymbol{F}_i) - \delta ln\left(A_i(\hat{\boldsymbol{F}}_i)\right), \tag{3.5}$$

where bold terms are vectors and damages are convex in mitigation if $\delta > 0$. This extends Bakkensen and Mendelsohn (2016) to allow for adaptation based on imperfect information. We also assume that adaptation is described by the linear relationship in Bakkensen and Mendelsohn (2016): $ln(A_i(\boldsymbol{F}_i)) = \gamma_1 ln(\boldsymbol{V}_i) + \gamma_2 ln(\boldsymbol{F}_i)$ and reparameterize (3.5) by adding and subtracting adaptation under perfect information to get

$$ln(d_i) = c + (\alpha - \delta\gamma_1) ln(\boldsymbol{V}_i) + (\beta - \delta\gamma_2) ln(\boldsymbol{F}_i) + \delta\gamma_2 \left[ln(\boldsymbol{F}_i) - ln\left(\hat{\boldsymbol{F}}_i\right)\right], \tag{3.6}$$

where the final term in (3.6) captures the distance between the actual and predicted intensity and/or location of the storm. This corresponds directly to the interpretation of (3.4) above where an optimistic bias in private beliefs feeds into ex-ante adaptation decisions: $P_i(F_i) - P_i(\hat{F}_i) \approx A_i(F_i) - A_i(\hat{F}_i)$ and then into higher damages.

This formulation does not account for uncertainty around the forecast. It can be thought of as a pure representation of the false sense of security hypothesis. To capture the joint impact of forecast accuracy and uncertainty around the forecast, we can replace the final term in (3.6) with a general measure of forecast uncertainty:

$$ln\,(d_i) = c + (\alpha - \eta\gamma_1) ln\,(\boldsymbol{V}_i) + (\beta - \eta\gamma_2) ln\,(\boldsymbol{F}_i) + \eta ln\left(U_i(\boldsymbol{F}_i, \hat{\boldsymbol{F}}_i)\right) + \epsilon_i, \tag{3.7}$$

where $\boldsymbol{U}_i\,(\cdot)$ jointly captures accuracy and uncertainty about or surrounding the forecast and ϵ_i indicates the residual or approximation error when going from (3.6) to (3.7).

There are many different measures of forecast uncertainty. A popular measure for point forecasts is the mean square forecast error (MSE); see Ericsson (2001). While the MSE produces a fixed measure over time, Jurado et al. (2015) propose a time-varying measure which combines MSE's across variables using a dynamic factor model with stochastic volatility. A popular measure for density forecasts is the log score (Mitchell and Wallis, 2011) which evaluates the predicted density, $\hat{g}_i\,(\cdot)$, at $\boldsymbol{F}_i$ conditional on the prediction $\hat{\boldsymbol{F}}_i$. When $\boldsymbol{F}_i$ falls in the tails of $\hat{g}_i\,(\cdot)$, it has a lower probability and so is associated with higher uncertainty. Another measure is the continuous ranked probability score (CRPS), which compares observations against the predicted cumulative distribution function. However, neither of these measures distinguish between $\boldsymbol{F}_i$ falling in the upper of lower tail of the distribution.

Rossi and Sekhposyan (2015) propose an alternative measure of 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}

$$U_{1,i}\left(F_{1i}, \hat{F}_{1i}\right) = \int_{-\infty}^{F_{1i}} \tilde{g}_i\left(x_1 | \hat{F}_{1i}\right) dx_1, \tag{3.8}$$

which captures how likely it is to observe F_i given the predicted distribution. The innovation is that $\tilde{g}_i\,(\cdot)$ is computed using historical forecast errors. An important distinction between (3.8) and Rossi and Sekhposyan (2015) is that $\tilde{g}_i\,(\cdot)$ is allowed to change across events (or time) so as to capture changes in the distribution (Hendry and Mizon, 2014). This ensures that $\tilde{g}_i\,(\cdot)$ is measured using information most relevant for adaptation decisions. So if F_{1i} is large relative to $\hat{F}_{1i}$ and the difference is big compared to past prediction errors, then there is more uncertainty than when F_{1i} is small relative to $\hat{F}_{1i}$. Thus, (3.8) captures both forecast accuracy as well as the uncertainty surrounding or associated with the forecast.

We can extend the measure to multiple variables and/or forecast horizons through the joint predictive distribution function. If we define $\boldsymbol{F}_i = \{F_{1i}, \ldots, F_{Ji}\}'$ for J potential variables and/or horizons, then the multivariate extension of (3.8) is

$$U_{J,i}\left(\boldsymbol{F}_i,\hat{\boldsymbol{F}}_i\right)=\int_{-\infty}^{F_{1i}}\ldots\int_{-\infty}^{F_{j,i}}\tilde{g}_i\left(\left\{x_{1i,},\ldots,x_{Ji}\right\}'\,|\hat{\boldsymbol{F}}_i\right)dx_1\ldots dx_J, \tag{3.9}$$

which captures both variable specific and common aspects of uncertainty across variables. This measure can be decomposed to illustrate what information would be lost if multiple variables / horizons actually mattered but only a single variable / horizon was considered:

$$ln\left(U_{J,i}(\boldsymbol{F}_i,\hat{\boldsymbol{F}}_i)\right)=ln\left(U_{1,i}(F_{1i},\hat{F}_{1i})\right)+ln\left(\int_{-\infty}^{F_{2i}}\ldots\int_{-\infty}^{F_{j,i}}\tilde{g}_i\left(\left\{x_{2i,},\ldots,x_{ji}\right\}'\,|F_{1i},\hat{\boldsymbol{F}}_i\right)dx_2\ldots dx_J\right), \tag{3.10}$$

where the second term in (3.9) represents the information loss. If the unobseved forecast errors are highly correlated with the observed error then the variation of the conditional density will be limited and so it is better to focus on the observed forecast errors.

Further simplification occurs if $\tilde{g}_i(\cdot)$ follows a normal distribution. We show in Technical Appendix B.1 that for small distances between F_i and $\hat{F}_i$ (3.8) can be approximated as

$$U_{1,i}(F_{1i},\hat{F}_{1i})\approx\frac{F_{1i}-\hat{F}_{1i}}{\hat{\sigma}_{i,(F_{1\_},\hat{F}_{1\_})}}, \tag{3.11}$$

where $\hat{\sigma}_{i,(F_{1\_},\hat{F}_{1\_})}$ is the 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. When (3.11) 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 measure is also related to comparisons of ex-post and ex-ante forecast uncertainty; see Clements (2014) and Rossi et al. (2017).

We can generalize this measure further by taking logs and relaxing the fixed 1-to-1 relationship between forecast accuracy and ex-ante risk

$$ln\left(U_{1,i}(F_{1i},\hat{F}_{1i})\right)\approx\frac{\eta_1}{\eta}\ln(\left|F_{1i}-\hat{F}_{1i}\right|)+\frac{\eta_2}{\eta}\ln(\hat{\sigma}_{i,(F_{1\_},\hat{F}_{1\_})}), \tag{3.12}$$

which allows us to assess their relative importance for damages.[2] In this context, forecast accuracy captures the errors in beliefs that individuals have about the location or severity of a disaster while ex-ante risk modulates the strength of those beliefs. Plugging (3.12) back into (3.7), allows us to test the hypothesis that forecast accuracy matters after accounting

[2]Note that while the forecast errors are constrained to be positive, an indicator function could be added to capture negative forecast errors.

for the ex-ante uncertainty. The next section discusses how to assess this relationship in the context of short-term forecasts and damages from hurricane strikes.

4 Short-term forecasts and Hurricane Damages

Tropical cyclones are powerful natural events that occur intermittently around the globe. They are an intrinsic part of the climate system in that they are fueled by large air-sea surface temperature differentials and play an important role in mixing different ocean layers to help distribute heat (Emanuel, 2001). Tropical cyclones are also among the most destructive climate events accounting for six of the top ten costliest global natural disasters since 1980 (MunichRe, 2018). Hurricanes, which are tropical cyclones that occur in the Atlantic and northeastern Pacific oceans, account for seven of the top ten costliest weather and natural disasters in the United States over the same period (NOAA NCEI, 2018). Therefore, they provide an important application on which to test the implications of our framework.

The damage model in (3.7) encompasses many of the existing models of hurricane damages. Emanuel (2005), Nordhaus (2010) and Strobl (2011) implicitly set $\eta \equiv 0$ and $\alpha \equiv 1$ to examine the relationship between damages and natural hazards. Others set $\eta \equiv 0$ to investigate the relationship between damages and vulnerabilities; see Kellenberg and Mobarak (2008) and Geiger et al. (2016). Bakkensen and Mendelsohn (2016) allow for $\eta \neq 0$ but implicitly assume $\hat{F}_i \equiv F_i$.

We are interested in understanding the relationship between damages and forecast accuracy and uncertainty. However, instead of imposing restrictions on the other determinants, we focus on the general framework in (3.7). This allows us to test the implications of the theoretical framework which controlling for alternative variables and explaining existing results. We then use model selection to simplify the model and understand which variables are the most important drivers of damages. Our approach is broadly defined within a general-to-specific modeling framework since we start with a general model and then simplify it while using our theoretical framework as a guide. The rest of this section describes the methods and data that we use to estimate the model.

4.1 Methods

The general-to-specific (*Gets*) modeling framework is described in detail by Campos et al. (2005) and Hendry and Doornik (2014). Recent developments, by Hendry et al. (2008), Castle et al. (2015), Hendry and Johansen (2015) and Pretis et al. (2016), illustrate its usefulness across a range of applications. *Gets* modeling provides a way to simultaneously summarize and extend the literature. This contrasts with the approach of focusing exclusively on individual determinants of damages. We describe it in the current context as follows.

First, we construct a general unrestricted model (GUM), which includes all potentially relevant (theory-based or otherwise) determinants of hurricane damages. It is loosely assumed that the residuals of the GUM are iid normally distributed.[3] It is also assumed that the GUM is potentially sparse and so nests the local data generating process (LDGP). Under these conditions, *Gets* consistently recovers the same model as if selection began from the LDGP. This helps ensure valid post-selection inference; see Chernozhukov et al. (2015). Thus, formulation of the GUM is an integral part of the process and requires effort to ensure that all potentially important drivers of hurricane damages are included.

Including a large number of explanatory variables can result in spurious correlations and misleading inference. *Gets* tackles this by simplifying the GUM based on the 'encompassing principle' (Mizon and Richard, 1986) so that each reduction exhibits minimal information loss based on a user-specified 'target gauge'. The target gauge plays the same role as regularization in other machine learning or model selection procedures (Mullainathan and Spiess, 2017) in that it seeks control the loss of information in the selection procedure. While regularization parameters are typically chosen empirically based on model performance, the target gauge is chosen beforehand and has a theoretical interpretation in that it determines the false-retention rate of variables in expectation, i.e. the 'gauge' (Castle et al., 2011 and Johansen and Nielsen, 2016). In practice, the target gauge is set based on the number of variables being selected over, so that *on average* a single irrelevant variable is kept.

The final model is chosen so that it provides a parsimonious explanation of the GUM conditional on the acceptable amount of information loss. This approach can be used to

[3]Hurricane damages approximate a log-normal distribution; see Willoughby (2012), Blackwell (2014) and Appendix Figure A.1.

search across many different model reduction paths in order to minimize potential path dependencies; see Hendry and Doornik (2014). If multiple models are retained, information criteria can be 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 methods available. We 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). We also assess the robustness of 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).

4.2 Data

Although data on hurricane strikes goes back to the 1850s, we focus on hurricane strikes in the Atlantic basin of the continental United States since 1955 for which a continuous database of hurricane forecasts exists. We start by describing the number of hurricane strikes and the sources for damages used in our analysis. Next, we describe the forecasts, the errors, and how we measure forecast uncertainty. Finally, in the remainder of this section we describe any additional variables used in the analysis.

The hurricane research division of the U.S. National Oceanic and Atmospheric Administration (NOAA) maintains a list of every storm with hurricane force winds to make landfall in the continental United States since 1851. 192 hurricanes made landfall in the Atlantic basin between 1900 and 2015.[4] Of these, 88 occurred between 1955 and 2015. Accounting for the fact that some hurricanes struck in 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 unique strikes between 1955 and 2015. Our analysis focuses on the damages for 98 of these strikes after removing cases for which forecasts are not available.

We collate damages for each strike from multiple sources. Damages are taken from annual Atlantic Hurricane Season reports (1955-2015) following Pielke Jr and Landsea (1998).[5]

[4]The hurricane research division maintains two lists of U.S. Atlantic landfalls. We use most up-to-date where the differences since 1955 are that it includes Helene [1958] and Ophelia [2005] but excludes Diane [1955].

[5]Reports were published in the Monthly Weather Review through 2011 and are available from the Hurricane

Table 4.1: Comparing Hurricane Damages by Source

Source	Obs	Median		Std. Dev.		Min		Max		Corr.	Similar
		dif	(%)	dif	(%)	dif	(%)	dif	(%)	(%)	(%)
Pielke Jr and Landsea (1998)	52	-	-	561	28	-3,225	-95	739	100	99.6	75.0
Pielke Jr et al. (2008)	79	-	-	3,907	322	-33,291	-95	5,533	2,186	99.2	64.6
ICAT	86	-	-	4,643	52	-33,291	-95	1,159	388	99.3	66.3
NOAA: Storm Events	81	-52	-27	11,860	46	-91,736	-96	2,117	180	88.2	9.9
NOAA: Billion-Dollar	27	373	4	4,713	28	-6,899	-32	20,961	117	99.2	7.4

Notes: All external sources are expressed relative to the current dataset. Dif is calculated as external minus current damages. All values are in millions of 2017 dollars. % dif is computed by dividing dif by current damages to get a percentage difference. A positive value implies that the external source has higher damages, whereas a negative number implies that current damages are higher. Corr. is the correlation between external and current damages. Similar is the share of observations for which the absolute percentage difference is less than or equal to 1 percent.

These are supplemented using individual tropical cyclone reports (1955-2015) and are updated using data from NOAA's hurricane research division; see Blake et al. (2011).[6]

While there are many datasets on hurricane damages, their values are not entirely reliable. Pielke Jr and Landsea (1998) (updated and extended by Pielke Jr et al., 2008 and the ICAT database) compile damages from 1900-2012. NOAA's 'Storm Events' database, which the SHELDUS database relies on, catalogs damages associated with each storm event at the U.S. county level since 1959.[7] However, Smith and Katz (2013) find that there is a tendency to underestimate the most damaging storms. As a result, NOAA established the 'Billion-Dollar' database which provides damages for climate and weather disasters that caused at least $1 billion in damage since 1980.

Our dataset is consistent with existing ones. Table 4.1 shows that it is closest to Pielke Jr et al. (2008) and the ICAT datasets. However, there are important differences. Damages are revised for several hurricanes, notably Celia [1970]. There are also some hurricanes for which the damages are lower. For example, Agnes [1972] initially struck Florida as a hurricane. It then weakened and later re-intensified into a tropical storm causing damage in Pennsylvania, New Jersey and New York. We only include damages associated with the initial hurricane strike whereas other datasets include all of the damages associated with the storm.

Research Division until 2011. The National Hurricane Center maintains the annual summaries since 2012.

[6] Available from the National Hurricane Center from 1958-2016 and NOAA from 1954-2005.

[7] We do not use the Sheldus database due to concerns of under reporting and because damages are allocated equally across counties for each storm. A common solution (see Davlasheridze et al., 2017) is to redistribute damages based on wind speed which would biases the analysis if favor of wind speed as an explanatory variable for damages. Addressing these issues is beyond the scope of our analysis and is left for future research.

Our dataset is also comparable with damages from the 'Storm Events' and 'Billion-Dollar' databases. Damages tend to be higher than the 'Storm Events' database, which suffers from under-reporting, but lower than the 'Billion-Dollar' event database. However they typically fall within the upper and lower confidence intervals.

Hurricane Forecasts, Errors and Uncertainty

Hurricane forecasts have a long history in the United States. The U.S. government has produced hurricane forecasts since at least the 1850's. These forecasts have changed dramatically with the advent of new methods and technologies, particularly through the use of satellite technology and supercomputers; see Shuman (1989), Sheets (1990) and Rappaport et al. (2009) for a history of these changes.

The National Hurricane Center (NHC) maintains all historical hurricane forecasts since it was establishment in 1954. The NHC's 'official' forecasts form the basis for hurricane watches, warnings and evacuation orders. They are also widely distributed to and used by news outlets. The forecasts are not based on a single model and should not be considered the same across storms. In fact, they are a combination of many different models and forecaster judgment; see Broad et al. (2007).

The forecasts are available every 6 hours for the entire history of a storm. However, not all time periods are relevant for our analysis. This especially true since hurricanes can be active for up to a month (Ginger [1971]) and transect the entire Atlantic ocean. In order to focus on the most relevant forecast for damages, we relabel each forecast in terms of the number hours it was made before landfall. We start by rounding the timing of each landfall to the closest point in a 6-hour window. Thus, if a hurricane made landfall at 16:00 UTC then it is rounded to 18:00 UTC. Next, we subtract the length of the forecast horizon (h) from the landfall time to get the time at which the h-hour-ahead 'landfall forecast' was generated. So for example, the 12-hour-ahead landfall forecast of a storm that made landfall at 18:00 UTC was generated at 6:00 UTC.

We focus on the 12-hour-ahead landfall track forecasts for several reasons. First, the U.S. Army Corps of Engineers (2004) and Milch et al. (2018) emphasize that individuals often wait until the last minute and focus on the track. Second, the short-horizon track forecasts are available for virtually every U.S. hurricane strike going back to 1955.[8] Third, the track is an integral part of the NHC's forecasts of hurricane intensity including rainfall (Kidder et al., 2005), wind speed (DeMaria et al., 2009), and storm surges (Resio et al., 2017). Thus, there is strong support for focusing on the 12-hour-ahead track forecasts.

Hurricane track forecast errors are computed differently from typical forecast errors. The track errors are calculated as the distance between two points on the surface of a spheroid (Vincenty, 1975) so as to account for the curvature of the earth.[9] Thus, the track error is purely a distance measure (i.e. absolute error) and does not have a directional interpretation.

This calculation is problematic if the direction of the error matters for damages. For example, forecasts that are too slow or biased to either side of the landfall location provide less warning time. Alternatively, forecasts that are too fast may underestimate the amount of rainfall and flooding and make individuals less likely to prepare for these damages. Despite these differences, each forecast error direction can be linked with the belief that the hurricane will be less destructive than it ultimately is.[10]

Panel A of Figure 4.1 plots the actual locations, 12-hour track forecasts and the difference between them for the closest available points to each hurricane strike, where the base of the arrow is the actual location and the head of the arrow is the projected location. The coloring is determined by the degree of the angle in terms of where the storm came from, where it actually is and where it was forecast to be with green indicating the storm moved faster than expected and red indicating it was slower than expected. While there is a mix of different types, there are more storms that were faster than expected.[11]

We can evaluate the the track forecast errors relative to the ex-ante risk associated with them at the time. To do this we estimate a time-varying measure of the historical forecast error density for each year using every forecast error at the 12-hour horizon in the previous five years.[12] As Appendix Figure A.3 shows, there have been large changes the estimated

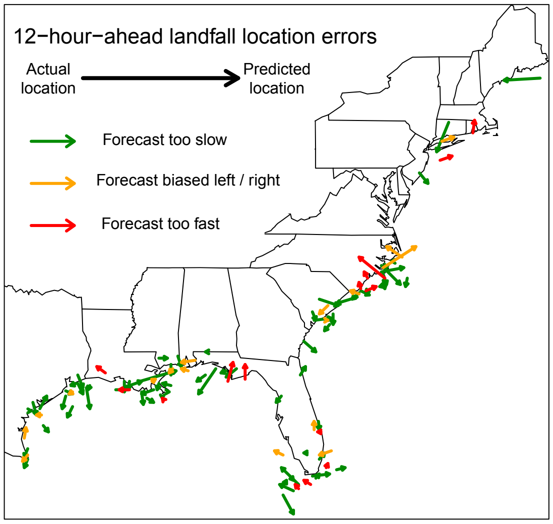
[8] No forecasts are available for Debra [1959] and Ethel [1960], likely due to their short duration. Longer horizon forecasts are available for even fewer storms and the intensity forecasts are only available since 1990.

[9] This calculation is more accurate than the more commonly used great-circle distance.

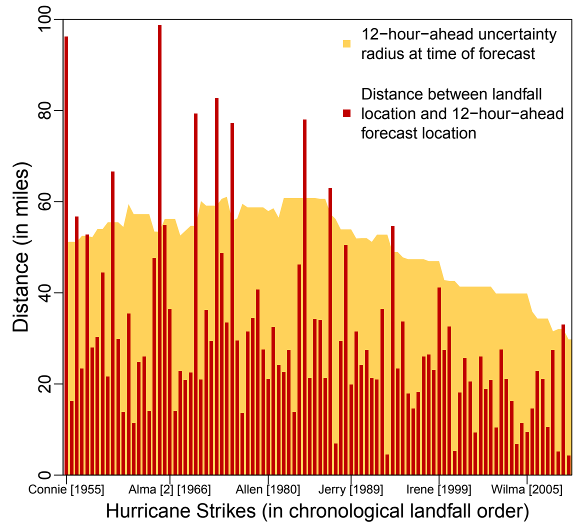
[10] By focusing on hurricane strikes we omit the costs of $p(\hat{F}) > p(F)$ when a hurricane did not strike.

[11] Robustness checks indicate that the results are unchanged after controlling for the size of the angle. However, the angle is positively associated with damages (not significant) which suggests that slower than expect storms are more damaging.

[12] Since the forecast database only extends back to 1954, the radius for storms prior to 1959 is estimated



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



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

Notes: Errors are computed using Vincenty (1975)'s formula for the distance between two points on the surface of a spheroid. 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 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 implied radius of uncertainty computed so that two-thirds of forecast errors for the five years prior to the strike fall within this area.

Figure 4.1: 12-Hour-Ahead Hurricane Landfall Track Errors

densities since 1955. Historically the densities have been skewed to the right with long tails driven by extreme outliers. However, since the 1990's, skewness has declined and a normal approximation is much more apt. The 66th percentile of these historical forecast error densities corresponds directly with how the NHC computes the radius of the 'cone of uncertainty', which has accompanied every hurricane track forecast since 2002. Therefore, using this distance provides a useful measure of the ex-ante risk associated with the forecast.

Panel B of Figure 4.1 plots the 12-hour-ahead landfall forecast errors along with our reconstructed measure of the 12–hour-ahead radius of uncertainty going back to 1955. The figure illustrates that landfall forecast errors have declined by around 60 percent over the past 60 years. The implied ex-ante risk has also declined over the same period so that both the errors and the expected risk associated with the forecasts has declined. However, the figure also shows that there are multiple hurricane strikes where the forecast errors at landfall exceeded their ex-ante risk; most recently for Sandy [2012].

using samples shorter than five years. Historically, the radius was also measured as the distance that captured all errors in the previous ten years. For more information see Broad et al. (2007) and `http://www.nhc.noaa.gov/aboutcone.shtml` (last accessed December 22, 2017).

Other Variables

While the focus of our analysis is on the link between damages and forecast uncertainty, we embed this relationship within a general model of damages. This ensures that we have a well specified model of hurricane damages. It also reduces the possibility that the relationship between forecast uncertainty and damages is confounded by omitted variable bias.

First, we include several measures of vulnerabilities. We use county level population and personal income estimates from the Bureau of Economic Analysis (BEA) since 1969. Prior to 1969, we 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. We 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. A similar approach is used for land area and housing units.[13]

Annual county level personal income prior to 1969 is estimated as follows. First, we assume that county level personal income shares were constant from 1955 to 1969. Second, we estimate annual income shares using a fixed effects panel data model and then starting in 1969, backcast the shares to 1955.[14] We combine the shares with state level income to get a county level estimate. We average these two approaches to get a robust measure.

We compute a 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, we aggregate across impacted counties. This alleviates concerns about county level estimates but requires us to choose which counties are impacted. We focus on coastal counties (Jarrell et al., 1992), which over-weights the importance of the coastal but is less likely to over-weight the impact of wind damages as in the approach used by Strobl (2011), Hsiang and Narita (2012), and Deryugina (2017).

[13] Pielke Jr et al. (2008) use a similar approach. We aggregate counties using BEA's modifications to Census codes: `https://www.bea.gov/regional/pdf/FIPSModifications.pdf` (last accessed November, 2016).

[14] 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 4.2: Data Descriptions and Summary Statistics

Variable	Description	Years	Min	Average	Max	Source
			Damages (D)			
DAMAGE	Nominal Damage (U.S. $1,000)	1955-2015	$28	$3,990,506	$105,900,000	NOAA, NHC
			Socio-economic Vulnerabilities (V)			
PD	Population Density (persons per acre)	1955-2015	12	257	3,940	BEA, Census
IP	Income Per Capita ($ 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
CRS	FEMA Community Rating System (rank)	1990-2015	7	9	10	FEMA
HMGP	Hazard Mitigation Grant Program (U.S. $1,000)	1990-2015	0	$38,042	$396,102	FEMA
			Natural Forces (F)			
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
			Forecast Uncertainty (U)			
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: Census Bureau; FEMA: Federal Emergency Management Agency; USACE: US Army Corps of Engineers; NLD: National Levee Database.

Next, we include measures of natural and climate forces. 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 obtained from the NHC. Maximum rainfall comes from NOAA's Weather Prediction Center.

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.[15] We then use the smoothed estimate for the month prior to the strike.

[15] We use the Hodrick-Prescott filter and set the smoothing parameter equal to 129,600 following Ravn and Uhlig (2002) for monthly data.

Finally, we compute storm-level estimates of sea surface air temperature following Estrada et al. (2015). The data are from NASA's global mean surface temperature index based on land-surface air temperature anomalies. The monthly series is then smoothed using the Hodrick-Prescott filter with a smoothing parameter equal to 129,600. The resulting estimate for the month prior to the hurricane strike is used. Table 4.2 and Appendix Figure A.2 provide sources, summary statistics, and plots of each of the variables.

5 A Model of Hurricane Damages

This section presents the general hurricane damages model followed by its estimation and reduction. The model includes all major 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:

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

The first line includes the vulnerabilities (V): housing unit density (hd), income per housing unit (ih) and 'real-time' hurricane strike location frequency (FREQ). We exclude population density since it is almost perfectly correlated with housing density and because housing has a more direct interpretation in this context.[16] The first two lines list the natural hazards (F): maximum rainfall (rain), storm surge (surge), negative minimum pressure (npress), maximum wind speed (wind), soil moisture relative to trend (MOIST), accumulated cyclone energy (ace) and global surface temperature (GST). The second line captures forecast accuracy and uncertainty (U): 12-hour-ahead forecast track errors (forc12) and the implied 12-hour-ahead radius of 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.

We estimate (5.1) using ordinary least squares (OLS). The estimated coefficients and their standard errors are reported in column (1) of Table 5.1. Several coefficients are statistically significantly different from zero. They include housing density, historical hurricane frequency, storm surge, central pressure, and the forecast errors. The coefficient on the forecast errors

[16] The results are similar when the population variables are included. Available upon request.

is positive, which is consistent with the prediction from the theoretical framework in (3.4) and (3.6). This supports the false sense of security hypothesis that larger forecast errors are associated with higher damages even after controlling for vulnerabilities and natural hazards.

Next we use model selection to discover the most important drivers of hurricane damages. While the significant variables in the full model provide an initial sense of this, model selection provides a more systematic approach. While we could impose restrictions to ensure that the selected model is consistent with our theoretical model, we start by selecting over all variables without any restrictions. Since the full model (i.e. the GUM) has 38 unique parameters (i.e. 37 variables plus the variance), then we set the target gauge equal to $\frac{1}{37} \approx 0.03$. We can adjust this target to understand how sensitive the results are to this choice. This helps address concerns that model selection is unstable (Mullainathan and Spiess, 2017).

There are 2^{37} ($>$ 130 billion) possible model combinations if we allow for every variable to be selected over. For a target gauge of 3 percent, the selection algorithm narrows the search space to 2^{16} ($<$ 70 thousand). In the process it eliminates entire branches of possible models and only estimates 335 candidate models. The algorithm finds that 9 terminal models are acceptable reductions of the GUM. The final model is selected from these terminal models using the Bayesian information criterion (BIC). It is also robust to alternative information criteria (see Appendix Table A.2).

Columns (2)-(4) in Table 5.1 present the selection results across a range of target gauges. Although the selected models are virtually identical, this masks large variation as the number of terminal models for each target range from $9 - 18$. In total, 48 unique terminal models are found across the different targets. Almost three quarters of these include some measure of forecast uncertainty, while almost 60 percent include the forecast errors themselves. For comparison, Bayesian model averaging (Zeugner and Feldkircher, 2015) suggests that the posterior inclusion probability of the forecast errors is between 40 and 60 percent depending on the choice of the prior distribution.[17]

The most important drivers of damages are housing and income, (-) central pressure, rainfall, storm surge, and the forecast errors. This is consistent with (3.7) in that the selected model has at least one measure of vulnerabilities (V), natural hazards (F), and

[17] Results available upon request.

Table 5.1: Damage Models by Selection Method

	(1) OLS	*Gets*			(5) Lasso OLS	(6) Double Lasso
Selection Target:		(2) 1%	(3) 3%	(4) 5%	BIC (10%)	BIC (10%)
Housing density	0.53**		0.40***	0.40***	0.45***	0.40**
	(0.24)		(0.15)	(0.15)	(0.15)	(0.16)
Income per housing unit	0.70	1.58***	1.28***	1.28***	1.05***	1.35***
	(0.79)	(0.14)	(0.18)	(0.18)	(0.17)	(0.47)
Historical fequency	-6.63**				-3.72**	-3.13
	(2.94)				(2.23)	(2.22)
Max rainfall	0.51	1.18***	1.14***	1.14***	0.74**	0.89**
	(0.43)	(0.32)	(0.31)	(0.31)	(0.32)	(0.32)
Max storm surge	1.21**	1.37***	1.34***	1.34***	1.30**	1.39***
	(0.48)	(0.39)	(0.37)	(0.37)	(0.37)	(0.38)
Min central pressure (-)	52.6***	52.9***	52.0***	52.0***	51.1***	49.0***
	(18.2)	(8.72)	(8.43)	(8.43)	(8.46)	(8.44)
Max wind speed	-0.27					
	(1.54)					
Soil moisture	0.77					
	(1.42)					
Seasonal cyclone energy	0.42				0.43*	0.59**
	(0.32)				(0.23)	(0.25)
Sea surface temperature	0.94					-0.25
	(3.21)					(1.68)
12-hour forecast errors	0.54*	0.55**	0.48**	0.48**		0.50**
	(0.29)	(0.24)	(0.23)	(0.23)		(0.24)
12-hour radius	3.17					0.82
	(2.60)					(1.37)
Trends:	Yes	No	No	No	No	No
Hour fixed effects:	Yes	No	No	No	No	No
Month fixed effects:	Yes	No	No	No	No	No
U.S. state fixed effects:	Yes	No	No	No	Yes	Yes
k	37	5	6	6	8	11
$\hat{\sigma}$	1.300	1.304	1.261	1.261	1.250	1.228
R^2	0.876	0.806	0.821	0.821	0.828	0.840
BIC	4.710	3.658	3.626	3.626	3.680	3.749

*p< 0.1 **p< 0.05 ***p< 0.01

Notes: Estimated using 98 observations including a constant and dummy variables for Gerda [1969] and Floyd [1987]. Standard errors are in parentheses. k is the number of selected regressors in the model.

forecast uncertainty (U) with each of their signs are in the expected direction. Selection is identical across a range of target gauges with the exception of housing density, which is not retained when the target gauge is set to 1 percent. This is in line with the full model except that rainfall and income are found to matter but historical hurricane frequency does not.

Wind speed, which is a common measure of natural hazards, does not appear in any of the selected models.[18] Instead, minimum central pressure is always included. This is supported by Bakkensen and Mendelsohn (2016), who find that central pressure provides a more reliable explanation of damages. In addition to central pressure, we also find that

[18]Even when it appears in the full GUM or one of the reduced GUM's, its coefficient has the wrong sign and is not significantly different from zero; see Appendix Table A.3.

rainfall and storm surge are empirically relevant. Rainfall leads to inland flooding damages whereas storm surge causes damages along the coast. Thus, we find a more disaggregated measure of natural hazards than previous models.

The results are fairly robust across alternative selection methods. We can perform model selection using Lasso, which shrinks the coefficients with a penalty (chosen here using BIC). We present the post-selection OLS coefficients (and standard errors), shown in column (5) of Table 5.1, to compare with the *Gets* results. More variables are retained, which is consistent with the fact that BIC roughly corresponds to a target gauge of 0.10 for the sample size and number of variables selected over; see Campos et al. (2003). The forecast errors are also not retained, but this is very sensitive to the choice of the regularization hyper-parameter.[19]

Regardless of which selection procedure is used, there is a concern that post-selection inference does not capture any uncertainty in the selection procedure. This can be addressed in several ways. Selection can be restricted so that it does not take place over the variables of interest; see Belloni et al. (2014), Hendry and Johansen (2015), and Chernozhukov et al. (2018). Alternatively, the standard errors (Van de Geer et al., 2014) or the critical values (Berk et al., 2013) can be adjusted to capture the additional uncertainty.

If the forecast errors are exogenous conditional on the other regressors, what we are interested in is the treatment effect of the forecast errors onto damages. Then we can perform 'Double Lasso' selection proposed by Belloni et al. (2014). This procedure is described in three steps. First, a Lasso regression is run on the full model of damages excluding the forecast errors. Second, another Lasso regression is run on a model of the forecast errors using all other explanatory variables. Third, the variables selected in the first two steps along with forecast errors are combined into a final OLS regression on damages. This allows for valid inference on the coefficient of the forecast errors as long as the underlying process is sparse. The results of this procedure are shown in the final column of Table 5.1 where the coefficient and standard errors corresponding to the landfall location forecast error are broadly consistent with OLS and the *Gets* model selection results.

Overall the results indicate that a small subset of drivers explain much of the variation in hurricane damages. The results provide further support for the false sense of security

[19] The forecast errors are included if cross-validation is used to determine the hyper-parameter.

hypothesis. The estimated coefficient on the forecast error variable is positive and significantly different from zero. This is in line with what the theoretical framework in (3.4) would predict. Furthermore, decomposing forecast uncertainty into forecast accuracy and ex-ante risk as in (3.12), illustrates that the forecast errors are included but not the ex-ante radius of uncertainty. This indicates that the forecast errors matter for hurricane damages.[20]

6 Robustness of the Results

This section evaluates the robustness of the results. We start by checking the robustness of the selected model to model misspecification. Next, we address potential concerns about omitted variable bias. Finally, we assess out-of-sample fit. Overall, these robustness checks confirm our finding that forecast errors matter for damages.

6.1 Model misspecification

Interpretation of the coefficients and their significance depends on whether the underlying assumptions about the model are satisfied. However, rerunning our analysis with a battery of diagnostic tests indicates that these assumptions may not be satisfied. The diagnostic tests shown in Table 6.1 are: the $\chi^2_{nd}(2)$ 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).[21] A rejection of the null hypothesis indicates that the assumption associated with that test is invalid. The diagnostic tests in column (1) indicate that the selected models have evidence of non-normal and heteroskedastic residuals. This provides evidence against the assumptions about the functional form and the log-linearity approximation.

Model misspecification can be dealt with in several ways. Heteroskedasticity of the residuals can be addressed by correcting the standard errors following Newey and West (1987). The model can also be extended by dropping outliers that may induce non-normality and adding squares of the explanatory variables to help capture possible nonlinearities that induce heteroskedasticity. These choices entail different trade-offs. Correcting the standard errors ensures consistent estimates without changing the model but does not address amy

[20]Ignoring any model uncertainty, then forecast errors are always significant across different model selection specifications, whereas ex-ante risk is not; see Appendix Table A.3.

[21]Note that since the data is irregularly spaced they are not really time series in a strict sense and so we do not report diagnostic tests for residual autocorrelation and time-varying variances.

underlying misspecification. Extending the model addresses misspecification but can make estimates less reliable if added variables are irrelevant.

Although the model can be expanded in different ways, we follow the approach advocated by Hendry and Johansen (2015). We embed the selected model into a more general model that also includes impulse dummies for every observation and squares of the variables. This manes that there are more variables than observations. However, selection is done over the impulse dummies and nonlinearities by exploiting the algorithm's ability to examine multiple block path searches. This is known as impulse indicator saturation (IIS); see Hendry et al. (2008) and Johansen and Nielsen (2016).[22] While this biases any further selection in favor of the model that was originally selected, it allows us to evaluate the robustness of the original selection results to model misspecification.

Castle et al. (2018) advocate for expanding and then selecting over the original GUM. In practice, this requires a tighter target to control the number of impulses retained and a looser target to capture marginally relevant variables. Hendry (2018) uses a two-step procedure where selection is done first over the impulse dummies with a tight target gauge and then over the entire model jointly with a looser target. This ensures that not too many impulses are retained without limiting the search space.

The selected model is broadly robust to misspecification. The heteroskedasticity corrected standard errors in column (2) of Table 6.1 do not indicate major changes in the significance of the coefficients. Extending the model produces similar results despite retaining the square of income and several impulse dummy variables; see column (3).[23] The nonlinear relationship between income and damages is in line with Geiger et al. (2016). The outlying observations capture several issues including measurement concerns in the late 1950s and the glancing landfalls by Alex [2004] and Arthur [2014]. No further diagnostic concerns are indicated when both impulses and nonlinearities are selected over jointly.[24] Importantly, the standard errors of the coefficients in column (3) are smaller than in column (1), which indicates that estimates are more reliable despite including more covariates.

[22]It is equivalent in this context to selecting over individual strike fixed effects.

[23]Selection is done using a target gauge of 1 percent but the results are identical when using a target gauge as tight as 0.01 percent.

[24]The normality test should be regarded with caution in this context; see Berenguer-Rico and Nielsen (2017).

Table 6.1: Robust Models of Hurricane Damages

	(1) *Gets* (3%)	(2) Gets+HCSE	(3) Extension (1%)	(4) Robust
Housing density	0.40***	0.40***	0.29***	0.27***
	(0.15)	(0.15)	(0.10)	(0.10)
Income per housing unit	1.28***	1.28***	1.40***	1.43***
	(0.18)	(0.21)	(0.13)	(0.15)
Min central pressure (-)	52.0***	52.0***	55.5***	56.6***
	(8.43)	(7.98)	(5.66)	(5.49)
Max rainfall	1.14***	1.14***	0.52**	0.57**
	(0.31)	(0.33)	(0.21)	(0.21)
Max storm surge	1.34***	1.34***	0.93***	0.99***
	(0.37)	(0.38)	(0.25)	(0.24)
12-hour forecast errors	0.48**	0.48*	0.30*	0.34**
	(0.23)	(0.27)	(0.16)	(0.15)
Income per housing unit sq.			0.45***	0.44***
			(0.09)	(0.09)
Outlying storms:				-3.44***
				(0.33)
Helene [1958]			-2.70***	
			(0.86)	
Cindy [1959]			-4.22***	
			(0.86)	
Gracie [1959]			-2.73***	
			(0.84)	
Alma [1] [1966]			-4.41***	
			(0.84)	
Bret [1999]			-2.48***	
			(0.87)	
Alex [2004]			-3.56***	
			(0.83)	
Arthur [2014]			-4.17***	
			(0.88)	
$\hat{\sigma}$	1.261	1.261	0.816	0.812
R^2	0.821	0.821	0.932	0.927
$\chi^2_{nd}(2)$	8.34**	8.34**	0.91	0.03
	[0.015]	[0.015]	[0.636]	[0.986]
F_{Het}	1.76*	1.76*	0.88	0.80
	[0.069]	[0.069]	[0.585]	[0.675]
F_{Het-X}	1.02	1.02	0.97	0.94
	[0.457]	[0.457]	[0.532]	[0.570]
$F_{RESET23}$	1.30	1.30	1.35	1.09
	[0.279]	[0.279]	[0.265]	[0.341]

*p< 0.1 **p< 0.05 ***p< 0.01

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 tail probabilty associated with the null hypothesis of each diagnostic test statistic is in square brackets. Column (2) shows the heteroskedasticity corrected standard errors. Income per housing unit is demeaned to facilitate interpretability of the coefficients.

Since the coefficients of the outlying storms are all negative and approximately a similar magnitude, we can test whether they are of equal magnitude. Under the null hypothesis of equal magnitude, the likelihood ratio has a statistic of 5.763 which we compare against a chi-square distribution with 6 degrees of freedom. The Wald test has a statistic of 0.960 which we compare against the F-distribution with $(6, 82)$ degrees of freedom. In both cases we fail to reject the null hypothesis using any reasonable critical value. Using a single dummy variable for the outlying storms results in a more parsimonious model, see column (4), with less uncertainty around the estimated coefficients (see Hendry and Santos, 2005).

The coefficient on the forecast errors remains positive and significant even after accounting for model misspecification. The estimate varies between 0.30 – 0.52 with the estimate in column (4) lying near the middle of this range. This result is also robust to additional checks. Similar results are obtained if we first select over impulses while forcing in the GUM and then selecting over the full model.[25] We also obtain similar results if we augment the full model with the joint outlier dummy variable and the square of income and then select. The results are also robust to alternative measures of normalized damages (Appendix Table A.4) and alternative measures of forecast uncertainty (Appendix Table A.5).

6.2 Controlling for potentially omitted variables

So far we have assumed that the forecast errors are at least weakly exogenous for hurricane damages. This is a safe assumption if the forecast errors are randomly determined (conditional on the other covariates) and unrelated to other potentially relevant variables that are excluded from the analysis. However, this may not be the case. For example, there are non-random changes in the forecast errors over time. To fully assess the robustness of the results, we need to account for the possibility that omitted variables could bias the relationship between damages and the forecasts.

The forecast errors may be correlated with storm dynamics. A volatile storm is more difficult to forecast and can cause more damage. For example, the rapid intensification of a hurricane just prior to landfall is difficult to forecast and is also associated with higher damages; see Kaplan et al. (2010). Furthermore, a slow moving storm with high amounts of rainfall can also be forecast poorly. For example, Harvey [2017] experienced rapid intensification in its initial buildup, slowed down as it made landfall, and then dumped record breaking amounts of rain leaving devastation in its wake. In either case, storm dynamics are associated with increased forecast errors and higher damages. Thus, the relationship between forecast errors and damages may be moderated by the storm's dynamics.

Forecast errors may also be correlated with longer-term adaptation and technological

[25] One difference is the inclusion of historical hurricane frequency, which remains significant until outliers are accounted for but does not alter the other results.

change. Adaptation often goes hand-in-hand with efforts to improve hurricane forecasts and could therefore be correlated with improved forecasts. There is evidence of this in column (2) of Table A.5 where the unrestricted ex-ante risk is positively correlated with hurricane damages. Since ex-ante risk is effectively a five-year moving average of past forecast errors, changes in it are driven less by short-term randomness and more by longer-term trends such as government expenditures and technological advancements. Therefore, the relationship between forecast errors and damages may also be moderated by these longer-term changes.

Given the existence of non-random relationships between the forecast errors and variables excluded from the model, it is important to assess whether controlling for them alters the results. While there is no concise measure of storm dynamics, we construct an instrument based on a measure of forecast skill, which is regularly used to assess the performance of hurricane and weather forecasts; see Cangialosi and Franklin (2016). Forecast skill is measured as the ratio of the official forecast errors to naïve forecast errors from a simple climatology and persistence model. Since naïve forecasts suffer from the same natural variability as the official forecast, this measure should remove the effect of storm dynamics from the official forecast errors. Naïve hurricane forecasts are available for all hurricanes starting in 1970. We also control for hurricane warning lead times to proxy for hurricane evacuations. This is measured as the number of hours warning was issued before a strike divided by the length of the coastline under warning.

It is more difficult to control for longer-term adaptation efforts. We include the normalized length of protective levees from the US Army Corps of Engineers' National Levee Database as a measure of how location-specific efforts have changed over time. We also include the radius of uncertainty as a general proxy of longer-term technological improvement. The maximum processing speed of NOAA's supercomputers was also considered.[26] Together, these measures should control for longer-term adaptation efforts.

Alternative measures of adaptation include the U.S. Federal Emergency Management Agency's (FEMA) Community Rating System (CRS) as well as its Hazard Mitigation Grant Program (HMGP). The CRS was created in 1990 as a part of the National Flood Insurance Program to incentivize flood damage mitigation using reductions in flood insurance premi-

[26] Results available upon request.

Table 6.2: Controlling for storm dynamics and adaptation efforts

	Dynamics		Dynamics & Adaptation	
	(1) 12-hours	(2) 36-hours	(3) 12-hours	(4) 36-hours
Housing density	0.26**	0.24**	0.18	0.08
	(0.13)	(0.11)	(0.13)	(0.10)
Income per housing unit	1.45***	1.30***	2.04***	2.20***
	(0.25)	(0.20)	(0.32)	(0.30)
Income per housing unit sq.	0.40**	0.53***	0.77***	1.02***
	(0.18)	(0.16)	(0.23)	(0.27)
Min central pressure (-)	55.1***	60.2***	54.5***	58.9***
	(7.11)	(7.46)	(6.69)	(6.76)
Max rainfall	0.45*	0.25	0.47*	0.28
	(0.26)	(0.23)	(0.24)	(0.21)
Max storm surge	1.10***	0.94***	1.10***	0.93***
	(0.29)	(0.28)	(0.27)	(0.31)
H-hour error (IV)	0.55**	0.23	0.49*	0.34
	(0.27)	(0.19)	(0.26)	(0.20)
Warnings	0.31*	0.23*	0.38**	0.36***
	(0.16)	(0.12)	(0.15)	(0.12)
H-hour radius			2.95**	2.44***
			(1.20)	(0.88)
Levee length			-3.07	-5.59***
			(2.34)	(2.34)
Outlying storms dummy	-3.55***	-3.86***	-3.11***	-3.38***
	(0.51)	(0.56)	(0.54)	(0.53)
$\hat{\sigma}$	0.764	0.785	0.721	0.717
R^2	0.928	0.925	0.939	0.939
	*p< 0.1 **p< 0.05 ***p< 0.01			

Notes: All equations are estimated using 65 observations and include a constant and a dummy variable for Floyd [1987]. The standard errors are in parentheses. Heteroskedasticity corrected standard errors are shown in columns (3) and (4).

ums. The HMGP was established in 1988 and includes grant funding for damage mitigation efforts, including improved warning systems; see Davlasheridze et al. (2017). This additional analysis can be found in Appendix Table A.6.

Table 6.2 shows that we continue to find a significant relationship between errors and damages with an elasticity around 0.5. This suggests that the findings are robust to concerns about omitted variables. However, the results do not hold for longer forecast horizons. Although the direction of the relationship remains consistent with our hypothesis, the estimated coefficient is smaller and not statistically significantly different from zero. This suggests that less attention is paid to longer forecast horizons which is consistent with the survey findings in Milch et al. (2018).

Table 6.3: Relative Out-of-Sample Model Performance

	(1) Bias	(2) RMSE	(3) MAPE
Gets	0.89	0.66	0.94
Lasso OLS	0.69	0.58	0.84
Double Lasso	1.20	0.98	1.18
Robust Gets	**0.20**	**0.20**	**0.64**
Nordhaus (2010)	1.07	0.99	6.47

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

6.3 Out-of-sample fit

Another concern is that we over fit the model in sample. Although *Gets* does not explicitly consider goodness of fit in its selection procedure, this is a common feature of model selection and machine learning techniques that do. We can check for potential over fit by evaluating the out-of-sample performance of the model. Since our estimation sample only includes hurricanes that made landfall through 2015, we can evaluate the model performance using hurricanes that made landfall in 2016 through 2018. This provides seven additional observations against which we can assess the performance.

We compare the out-of-sample performance across several models. The first is the selected model using *Gets* from column (4) of Table 5.1. The second is the selected model using Lasso from column (5) of Table 5.1. The third is the selected model using Double Lasso from column (6) of Table 5.1. The fourth is the robust model from column (1) of Table A.5. The fifth is a simple model of damage based on the estimated relationship with income and central pressure (see: Bakkensen and Mendelsohn, 2016). The sixth, and final model, is a simple model of damage using a fixed relationship with income and an estimated relationship with central pressure (see: Nordhaus, 2010; Strobl, 2011).

The results are presented in Table 6.3. Overall, the robust model performs considerably better than any of the other models. This is in part driven by Harvey [2017] where the robust model almost perfectly predicts official damages. However, as Appendix Figure A.4 shows, the robust model also performs well for the other hurricanes. In fact, it never does worse than most models and official damages always fall within 1 standard deviation of the forecast. This reinforces the idea that the results are robust both in sample and out.

7 The value of improving hurricane forecasts

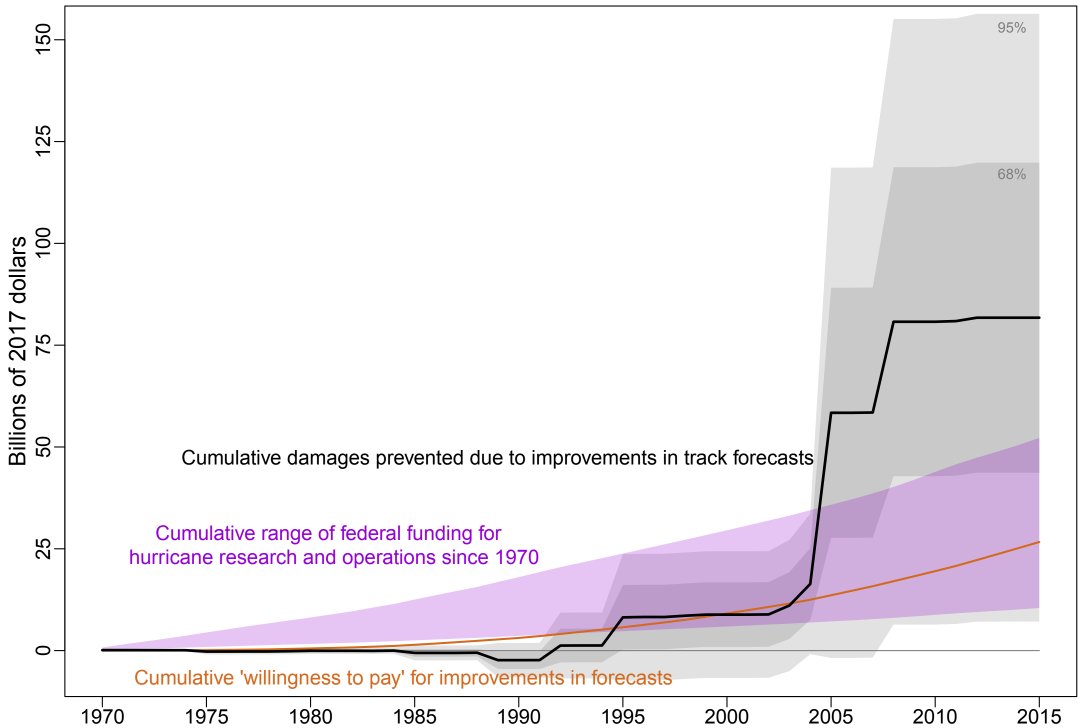
Next we seek to assess the implications of the findings given the consistency of the results. Assuming that the model is correctly specified and forecast errors are super exogenous (see Engle et al., 1983), then we can conduct a counterfactual experiment to assess the short-term impact of improvements in forecast accuracy on hurricane damages. There are several tests of super exogeneity by Engle and Hendry (1993), Hendry and Santos (2010), and Castle et al. (2017). Using the approach by Hendry and Santos (2010), we fail to reject the null hypothesis of super exogeneity.[27] Given that super exogeneity is satisfied, we can start by predicting what damages would have been using the average landfall-forecast errors from 1955-1969 for all strikes since 1970. From this we subtract damages predicted using actual forecast errors from each strike to get a prediction of the damages prevented since 1970.[28]

Figure 7.1 illustrates that our prediction of the cumulative damages prevented due to forecast improvements since 1970 is around \$82 billion. To put this into context, this value is 14 percent of the total predicted damages from 1970-2015 and is just shy of the damages caused by Hurricane Maria [2018]. We can conduct a cost benefit analysis by comparing the total benefit against the cumulative cost of producing the forecasts and their improvements since 1970, which were obtained from historical reports of the Office of the Federal Coordinator for Meteorology. This comparison illustrates that after accounting for the costs, the predicted net savings is around \$30 – 71 billion.

We can compare the prediction of damages prevented against the cumulative private willingness to pay for forecast improvements since 1970 as extrapolated from the findings of Katz and Lazo (2011). Figure 7.1 illustrates that damages prevented due to forecast improvements are greater than both public and private willingness to pay. This suggests that both individuals and the federal government have severely underestimated the value of improving hurricane forecasts. This is robust to alternative model specifications but should be considered a lower bound of the total net benefit from hurricane forecast improvements since we do not account for fatalities prevented (Willoughby et al., 2007) or reduced evacuation and damage mitigation costs (Regnier, 2008).

[27] Results available upon request.

[28] See Technical Appendix B.2 for details.



Notes: Damages prevented is calculated as the difference in damages estimated using the actual forecast error vs. damages estimated using the average forecast error from 1955-1969. The model used for estimation is column (2) of Table A.5. The 5% - 95% confidence interval around this estimate is computed using the delta method assuming independence and normality. See Technical Appendix B.2 for details. Federal funding for hurricane research and operations is taken from historical editions of the Office of the Federal Coordinator for Meteorology's Federal Plan for Meteorological Services and Supporting Research. The range is computed as being 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 7.1: The cost and benefits of improving forecast accuracy since 1970

8 Conclusions

Forecasts of natural disasters can alter the destructiveness of t event when operating through individuals' beliefs about the costs of damage mitigation. This is particularly true if individuals have a false sense of security about the accuracy of these imperfect forecasts. In this paper we test for and quantify this relationship using an empirical model of damages for all hurricanes to strike the continental United States in the past 60 years. We start by estimating the full empirical model using OLS. Next, we simplify the model using model selection methods and show that a small subset of drivers, including the 12-hour-ahead forecast errors, explain most of the variation in hurricane damages.

There is a positive and statistically significant relationship between the 12-hour-ahead landfall-forecast errors and damages. This relationship is consistent with the predictions of the false sense of security hypothesis. It is robust to outliers, alternative measures of uncertainty, model specifications, out-of-sample storms, and controls for storm dynamics and technological change. A one standard deviation increase in the distance between the storm's predicted and actual landfall location leads to $3,000 additional dollars in damages per household. Interpreting this result through the lens of the theoretical framework, it indicates that the forecasts play a critical role in guiding individuals' beliefs about the value of short-term damage mitigation efforts. It also illustrates that, in aggregate, individual decisions to protect and relocate property in the face of a disaster can have a significant impact on the overall cost of a natural disaster.

Focusing on the specific implications of the results, we find 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 other longer-term adaptation efforts have kept damages from rising faster than they otherwise would have. Comparing the cumulative damages prevented against the cost of producing the forecasts, we find that there is a net benefit of around $30 – 71 billion. This illustrates that improvements in hurricane forecasts over the past few decades produced benefits beyond the well-documented reduction in fatalities and have outweighed the associated costs.

This is particularly important since hurricanes are expected to become even more difficult to forecast in the future. Knutson et al. (2010) argue that climate change will increase hurricane intensity. As a result, in the future we are more likely to see hurricanes akin to Harvey [2017] whose storm dynamics are harder to predict (Emanuel, 2017a,b). In light of this reality, our findings support maintained investment in and continued measures to improve hurricane forecasting capabilities along with other longer-term adaptation efforts so that any future loss of life and property can be minimized.

Bibliography

Baker, S. R. and N. Bloom (2013). Does uncertainty reduce growth? using disasters as natural experiments. NBER Working Paper No. 19475, National Bureau of Economic Research.

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.

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.

Berenguer-Rico, V. and B. Nielsen (2017). Marked and weighted empirical processes of residuals with applications to robust regrssions. Department of Economics Disscusion Paper No. 841, University of Oxford.

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.

Blackwell, C. (2014). Power law or lognormal? Distribution of normalized hurricane damages in the United States, 1900–2005. *Natural Hazards Review 16*(3), 04014024.

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.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Bloom, N. (2014). Fluctuations in uncertainty. *The Journal of Economic Perspectives 28*(2), 153–175.

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., N. R. Ericsson, and D. F. Hendry (2005). General-to-Specific Modeling: An overview and selected bibliography. International Finance Discussion Paper No. 838, Board of Governors of the Federal Reserve System.

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.

Carsell, K. M., N. D. Pingel, and D. T. Ford (2004). Quantifying the benefit of a flood warning system. *Natural Hazards Review 5*(3), 131–140.

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 (2018). Robust model selection. *mimeo*.

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. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Chernozhukov, V., C. Hansen, and M. Spindler (2015). Valid post-selection and post-regularization inference: An elementary, general approach. *Annual Review of Economics 7*, 649–88.

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. (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. (2001). Contribution of tropical cyclones to meridional heat transport by the oceans. *Journal of Geophysical Research: Atmospheres 106*(D14), 14771–14781.

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 Satistical 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. (2018). First-in, First-out: Modelling the UK's CO_2 Emissions, 1960-2016. *mimeo*.

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 (2005). Regression models with data-based indicator variables. *Oxford Bulletin of Economics and statistics 67*(5), 571–595.

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.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), 263–292.

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.

Knight, F. H. (1921). *Risk, uncertainty and profit.* Houghton Mifflin.

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.

Krzysztofowicz, R. and D. R. Davis (1983). A methodology for evaluation of flood forecast-response systems. *Water Resources Research 19*(6), 1423–1454.

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.

Mizon, G. E. and J.-F. Richard (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica 54*(3), 657–678.

Mullainathan, S. and J. Spiess (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives 31*(2), 87–106.

MunichRe (2018). NatCatSERVICE. `https://natcatservice.munichre.com/`. [last accessed 27-September-2018].

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.

Neumayer, E. and F. Barthel (2011). Normalizing economic loss from natural disasters: a global analysis. *Global Environmental Change 21*(1), 13–24.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

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.

Pappenberger, F., H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, and J. Thielen (2015). The monetary benefit of early flood warnings in europe. *Environmental Science & Policy 51*, 278–291.

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 nowcast 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.

Sheets, R. C. (1990). The National Hurricane Center-past, present, and future. *Weather and Forecasting 5*(2), 185–232.

Shuman, F. G. (1989). History of numerical weather prediction at the National Meteorological Center. *Weather and Forecasting 4*(3), 286–296.

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.

Von Neumann, J. and O. Morgenstern (1945). *Theory of games and economic behavior.* Princeton University Press Princeton, NJ.

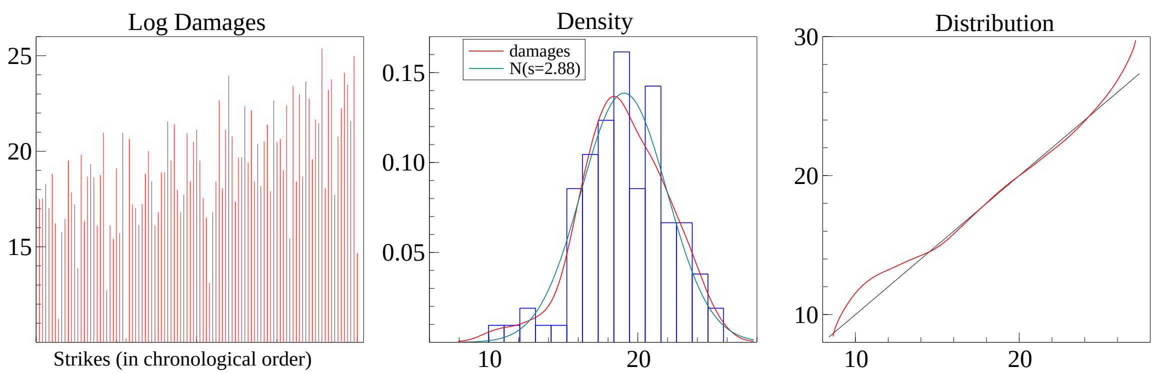
White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Willoughby, H. E. (2012). Distributions and trends of death and destruction from hurricanes in the United States, 1900–2008. *Natural Hazards Review 13*(1), 57–64.

Willoughby, H. E., E. Rappaport, and F. Marks (2007). Hurricane forecasting: The state of the art. *Natural Hazards Review 8*(3), 45–49.

Zeugner, S. and M. Feldkircher (2015). Bayesian model averaging employing fixed and flexible priors: The bms package for r. *Journal of Statistical Software 68*(4), 1–37.

A Appendix "For Online Publication"



Notes: Hurricane damages are the logs of the nominal values of hurricane damages for each hurricane strike.

Figure A.1: Hurricane Damages

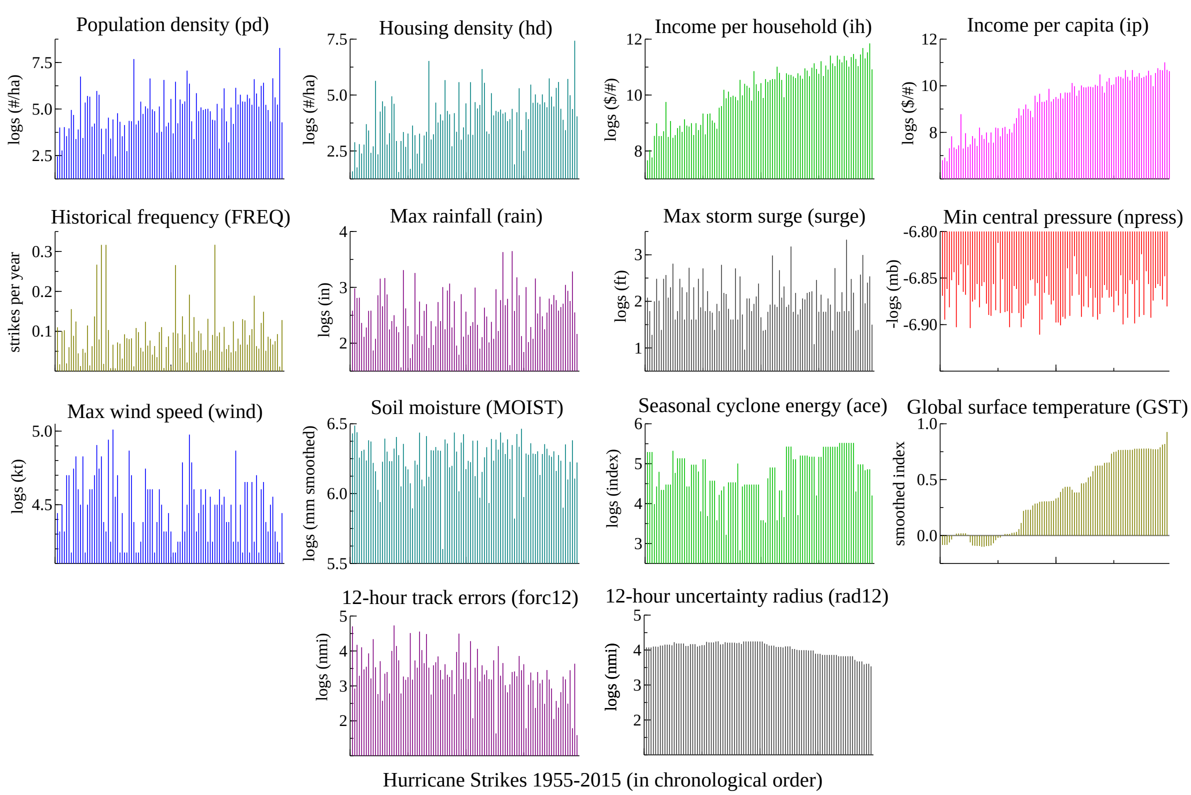
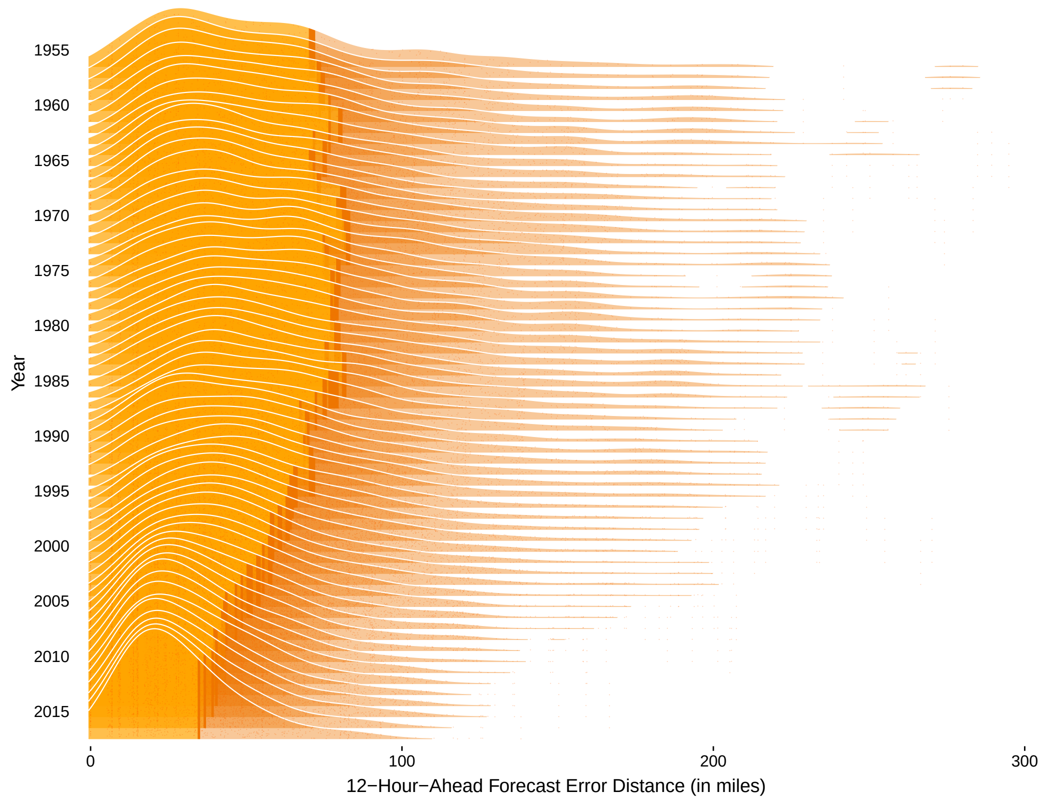


Figure A.2: Data Plots



Notes: Each density is estimated using all 12-hour forecast errors in the Atlantic Basin over the previous five years based on between 100-2000 observations. The orange bars represent the 67th percentile for each year.

Figure A.3: 12-Hour-Ahead Forecast Error Densities by Year

Table A.1: Bivariate Correlations

	pd	hd	ih	ip	FREQ	rain	surge	npress	wind	MOIST	ace	GST	forc12	cone12	Trend
dam	**0.51**	**0.51**	**0.59**	**0.54**	-0.14	0.32	**0.54**	**0.60**	0.34	0.09	0.30	**0.52**	-0.25	-0.43	**0.53**
pd		**0.98**	**0.55**	0.49	-0.02	0.14	0.12	0.11	-0.01	-0.02	0.05	0.39	-0.15	-0.29	0.41
hd			**0.64**	**0.61**	0.01	0.11	0.08	0.09	-0.05	-0.01	0.08	0.50	-0.19	-0.36	**0.53**
ih				**0.99**	0.02	0.11	0.07	0.05	-0.15	0.00	0.21	**0.93**	-0.42	**-0.63**	**0.95**
ip					0.04	0.07	0.03	0.02	-0.18	0.01	0.21	**0.94**	-0.42	**-0.64**	**0.97**
FREQ						-0.17	-0.10	-0.04	0.12	0.05	0.03	0.01	-0.04	-0.03	0.02
rain							0.09	0.11	0.01	0.22	0.19	0.14	-0.24	-0.20	0.12
surge								**0.65**	**0.52**	0.11	0.08	0.01	-0.10	-0.09	0.01
npress									**0.81**	-0.02	0.18	0.03	-0.01	-0.14	0.01
wind										0.03	0.12	-0.20	0.08	0.08	-0.19
MOIST											0.08	0.01	0.01	-0.02	-0.02
ace												0.35	-0.26	-0.43	0.28
GST													-0.46	**-0.79**	**0.98**
forc12														0.43	-0.44
cone12															**0.74**

Table A.2: Estimated Terminal Models for a target gauge of 3%

Terminal Model:	Final GUM	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Housing density	0.60***	0.58***	0.64***	0.61***	0.59***	0.40***	0.64***	0.56***	0.44***	0.55***
	(0.18)	(0.14)	(0.19)	(0.14)	(0.14)	(0.15)	(0.18)	(0.14)	(0.15)	(0.19)
Income per household	0.47		1.01***			1.28***	1.01***		1.22***	1.12***
	(0.58)		(0.19)			(0.18)	(0.17)		(0.20)	(0.21)
Historical fequency	-4.46*		-6.35**	-4.85**	-5.18**		-6.54***	-4.21*		
	(2.48)		(2.53)	(2.24)	(2.26)		(2.31)	(2.25)		
Max rainfall	0.59*	1.12***				1.14***		0.72**	0.97***	0.92***
	(0.36)	(0.32)				(0.31)		(0.32)	(0.31)	(0.34)
Max storm surge	1.16***	1.44***		1.41***	1.45***	1.34***	1.30***	1.24***	1.30***	1.27***
	(0.41)	(0.38)		(0.38)	(0.39)	(0.37)	(0.38)	(0.39)	(0.38)	(0.38)
Min central pressure (-)	63.0***	53.9***	87.7***	53.6***	53.2***	52.0***	52.8***	51.2***	52.5***	68.3***
	(14.0)	(8.66)	(13.0)	(8.73)	(8.90)	(8.43)	(8.58)	(8.62)	(8.57)	(13.9)
Max wind speed	-1.03		-1.59							-1.53
	(1.16)		(1.23)							(1.17)
Seasonal cyclone energy	0.53**			0.57**	0.61**		0.47*	0.42*	0.48*	
	(0.25)			(0.25)	(0.26)		(0.24)	(0.24)	(0.25)	
Soil Moisture	0.85		1.69*							
	(0.97)		(0.95)							
12-hour forecast error	0.52**	0.44*			0.33	0.48**		0.53**		
	(0.24)	(0.24)			(0.24)	(0.23)		(0.24)		
12-hour radius	1.95	2.35*		2.96**	2.52**				1.23	0.88
	(1.44)	(1.19)		(1.23)	(1.25)				(1.01)	(0.99)
Strike trend	0.03	0.05***		0.05***	0.05***			0.04***		
	(0.02)	(0.01)		(0.01)	(0.01)			(0.01)		
AUG	0.20		0.16							0.15
	(0.29)		(0.32)							(0.301)
NY	-1.05		-1.78**				-1.54**			-0.85
	(0.85)		(0.81)				(0.74)			(0.80)
VA	1.73**			1.55*				1.79**		
	(0.82)			(0.78)				(0.81)		
NC	-0.71*		-0.74**					-0.74**		
	(0.42)		(0.35)					(0.36)		
k	16	7	9	8	8	6	7	10	7	9
Log-likelihood (-)	147.9	157.6	161.3	157.6	158.8	157.1	157.5	153.5	157.1	157.8
AIC	3.406	3.420	3.537	3.441	3.464	**3.389**	3.417	3.397	3.411	3.464
HQ	3.609	3.527	3.666	3.558	3.582	**3.485**	3.524	3.536	3.517	3.592
BIC	3.908	3.684	3.854	3.731	3.754	**3.626**	3.681	3.740	3.675	3.780
$\hat{\sigma}$	1.219	1.275	1.340	1.282	1.298	1.261	1.273	1.244	1.269	1.291
R^2	0.851	0.819	0.804	0.819	0.814	0.821	0.819	0.833	0.821	0.818

*p< 0.1 **p< 0.05 ***p< 0.01

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 target gauge (target value) 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 in the model.

Table A.3: Final Estimated GUMs at different targets

Target gauge:	(Original GUM)	(All)	1%	2%	3%	4%	5%
Housing density	0.53**	0.50**	0.46**	0.43**	0.60***	0.44**	0.46**
	(0.24)	(0.21)	(0.20)	(0.20)	(0.18)	(0.20)	(0.20)
Income per household	0.70	0.74	0.75	1.08	0.47	0.79	0.87
	(0.79)	(0.70)	(0.68)	(0.67)	(0.58)	(0.68)	(0.61)
Historical fequency	-6.63**	-6.81**	-7.09***	-6.14**	-4.46*	-6.30**	-6.10**
	(2.94)	(2.70)	(2.58)	(2.68)	(2.48)	(2.63)	(2.57)
Max rainfall	0.51	0.59	0.59	0.67*	0.59*	0.61	0.65*
	(0.43)	(0.38)	(0.37)	(0.38)	(0.36)	(0.37)	(0.36)
Max storm surge	1.21**	1.13**	1.15***	1.18***	1.16***	1.26***	1.25***
	(0.48)	(0.43)	(0.42)	(0.43)	(0.41)	(0.41)	(0.40)
Min central pressure (-)	52.6***	56.3***	51.8***	59.5***	63.0***	53.6***	53.4***
	(18.2)	(15.0)	(9.25)	(14.8)	(14.0)	(14.4)	(13.9)
Max wind speed	-0.27	-0.49		-0.98	-1.03	-0.49	-0.65
	(1.54)	(1.25)		(1.24)	(1.16)	(1.23)	(1.17)
Seasonal cyclone energy	0.42	0.32	0.31	0.45	0.53**	0.38	0.38
	(0.32)	(0.28)	(0.28)	(0.27)	(0.25)	(0.27)	(0.25)
Soil moisture	0.77	0.57	0.65	0.11	0.85		
	(1.42)	(1.09)	(1.03)	(1.05)	(0.97)		
Sea surface temperature	0.94	0.13	0.23	-1.79		0.18	
	(3.21)	(2.51)	(2.39)	(2.28)		(2.47)	
12-hour forecast error	0.54*	0.62**	0.62**	0.59**	0.52**	0.64**	0.69***
	(0.29)	(0.26)	(0.25)	(0.26)	(0.24)	(0.25)	(0.24)
12-hour radius	3.17	2.48	2.55	0.51	1.95	2.43	2.11
	(2.60)	(2.06)	(2.01)	(1.65)	(1.44)	(1.95)	(1.62)
Trends:	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hour fixed effects:	Yes	Yes	Yes	Yes	No	Yes	No
Month fixed effects:	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S. State fixed effects:	Yes	Yes	Yes	Yes	Yes	Yes	Yes
k	37	27	24	22	16	24	21
$\hat{\sigma}$	1.300	1.234	1.210	1.247	1.219	1.219	1.203
R^2	0.876	0.869	0.868	0.856	0.851	0.866	0.864

*p< 0.1 **p< 0.05 ***p< 0.01

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. 'All' combines the retained variables from each Final GUM. k is the number of selected regressors in the model.

Table A.4: Alternative Measures of Normalized Damages

	(1) Nominal	(2) Real	(3) Norm-1	(4) Norm-2	(5) Norm-3
Housing density	0.27***	0.35***	0.43***	0.38***	-0.78***
	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)
Income per housing unit	1.43***	0.74***	-0.00	-0.01	0.27**
	(0.12)	(0.19)	(0.13)	(0.14)	(0.13)
Income per housing unit sq.	0.44***	0.42***	0.38***	0.42***	0.38***
	(0.09)	(0.09)	(0.10)	(0.10)	(0.09)
Min central pressure (-)	56.6***	57.2***	61.2***	61.2***	46.1***
	(5.49)	(5.53)	(6.06)	(6.11)	(5.82)
Max rainfall	0.57***	0.56***	0.66***	0.64***	0.35
	(0.21)	(0.21)	(0.23)	(0.23)	(0.22)
Max storm surge	0.99***	1.06***	0.96***	0.95***	0.98***
	(0.24)	(0.25)	(0.27)	(0.27)	(0.26)
12-hour forecast errors	0.34**	0.33**	0.24	0.29*	0.40**
	(0.15)	(0.15)	(0.17)	(0.17)	(0.16)
Outlying storms dummy	-3.44***	-3.47***	-3.63***	-3.56***	-3.20***
	(0.33)	(0.34)	(0.37)	(0.37)	(0.35)
$\hat{\sigma}$	0.812	0.817	0.896	0.902	0.860
R^2	0.927	0.908	0.882	0.879	0.853

*p< 0.1 **p< 0.05 ***p< 0.01

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. The different normalizations are: (2) CPI inflation; (3) Pielke Jr and Landsea (1998); (4) Pielke Jr et al. (2008); (5) Neumayer and Barthel (2011)

Table A.5: Alternative Measures of Uncertainty

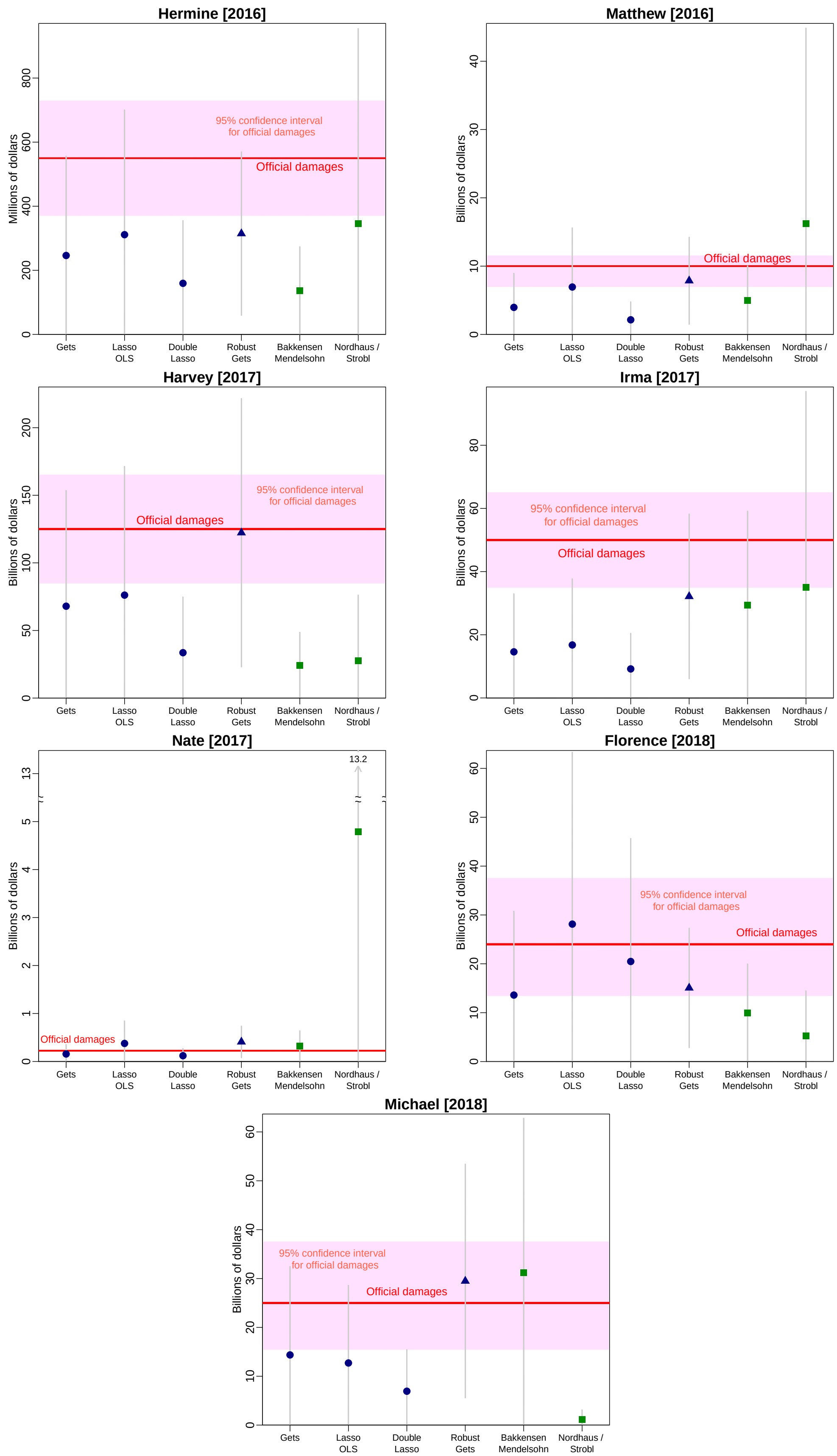
	(1) Errors	(2) Errors & Radius	(3) Errors/Radius	(4) R&S(2015)
Housing density	0.27***	0.24**	0.28***	0.28***
	(0.10)	(0.09)	(0.10)	(0.10)
Income per housing unit	1.43***	1.76***	1.36***	1.36***
	(0.12)	(0.19)	(0.12)	(0.12)
Income per housing unit sq.	0.44***	0.66***	0.41***	0.41***
	(0.09)	(0.13)	(0.09)	(0.09)
Min central pressure (-)	56.6***	57.8***	56.8***	56.9***
	(5.49)	(5.41)	(5.55)	(5.58)
Max rainfall	0.57***	0.56***	0.55***	0.54**
	(0.21)	(0.20)	(0.21)	(0.21)
Max storm surge	0.99***	0.98***	0.97***	0.97***
	(0.24)	(0.24)	(0.25)	(0.25)
12-hour forecast errors	0.34**	0.27*		
	(0.15)	(0.15)		
12-hour radius		2.10**		
		(0.96)		
12-hour error/radius			0.29*	
			(0.16)	
12-hour uncertainty				0.23*
				(0.14)
Outlying storms dummy	-3.44***	-3.19***	-3.49***	-3.50***
	(0.33)	(0.35)	(0.34)	(0.34)
$\hat{\sigma}$	0.812	0.795	0.819	0.822
R^2	0.927	0.930	0.925	0.925
*p< 0.1 **p< 0.05 ***p< 0.01				

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)

Table A.6: Controlling for Storm Dynamics and Adaptation Efforts

	(1) Final	(2) (1)+ RAD+LEV	(3) (1)+ CRS+HMG	(4) (1)+ (2)+(3)	(5) (1)+ Naïve	(6) (5)+ RAD+LEV	(7) (5)+ CRS+HMG	(8) (5)+ (6)+(7)
Housing density	0.27	0.16	0.17	0.10	0.25	0.15	0.15	0.09
	(0.18)	(0.19)	(0.21)	(0.22)	(0.18)	(0.19)	(0.21)	(0.22)
Income per housing unit	4.12***	4.39***	4.17***	4.24***	4.18***	4.41***	4.23***	4.23***
	(1.16)	(1.16)	(1.23)	(1.23)	(1.17)	(1.18)	(1.24)	(1.24)
Income per housing unit sq.	-1.07	-0.64	-1.01	-0.51	-1.14*	-0.70	-1.07	-0.57
	(0.62)	(0.66)	(0.64)	(0.70)	(0.63)	(0.67)	(0.65)	(0.71)
Min central pressure (-)	65.6***	64.5***	63.8***	63.2***	64.7***	63.7***	62.9***	62.4***
	(9.43)	(9.32)	(9.77)	(9.67)	(9.55)	(9.47)	(9.91)	(9.82)
Max rainfall	0.20	0.16	0.26	0.19	0.21	0.17	0.27	0.19
	(0.31)	(0.31)	(0.32)	(0.32)	(0.31)	(0.31)	(0.32)	(0.33)
Max storm surge	1.09**	1.03**	1.15**	1.12**	1.09**	1.03**	1.16**	1.12**
	(0.40)	(0.40)	(0.42)	(0.42)	(0.40)	(0.40)	(0.42)	(0.42)
12-hour official error	0.39	0.29	0.41	0.36	0.37	0.28	0.39	0.35
	(0.23)	(0.24)	(0.25)	(0.25)	(0.24)	(0.24)	(0.26)	(0.26)
12-hour naïve error					0.19	0.17	0.19	0.19
					(0.24)	(0.26)	(0.25)	(0.262)
12-hour radius		2.58		2.76		2.53		2.77
		(1.52)		(1.72)		(1.54)		(1.74)
Levee length		-1.61		-1.74		-0.79		-0.85
		(3.17)		(3.25)		(3.42)		(3.51)
Community rating system			-0.17	-0.19			-0.18	-0.20
			(0.22)	(0.21)			(0.22)	(0.22)
HMG spending per capita			-0.02	0.00			-0.02	0.00
			(0.03)	(0.03)			(0.03)	(0.04)
Outlying storms dummy	-3.56***	-3.22***	-3.57***	-3.15***	-3.58***	-3.23***	-3.59***	-3.14***
	(0.55)	(0.59)	(0.57)	(0.64)	(0.56)	(0.60)	(0.58)	(0.65)
$\hat{\sigma}$	0.783	0.772	0.797	0.788	0.788	0.780	0.802	0.795
R^2	0.923	0.930	0.925	0.932	0.924	0.931	0.927	0.933
	*p< 0.1 **p< 0.05 ***p< 0.01							

Notes: All equations are estimated using 41 observations and include a constant. The standard errors are in parentheses.



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

Figure A.4: Out-of-sample damage 'predictions' by method and storm

B Technical Appendix

B.1 Approximation of the Normal Distribution

Given the assumption that the past forecast errors have a normal density then

$$g_i(F_{1i}|\hat{F}_{1i}) = \frac{1}{\sqrt{2\pi}\sigma_{i,(F_{1\_},\hat{F}_{1\_})}} e^{-(F_{1i}-\hat{F}_{1i})^2/2\sigma_i^2}, \tag{B.1}$$

which we can rewrite in terms of a standard normal density as

$$g\left(z_i\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z_i{}^2}{2}}, \tag{B.2}$$

where $z_i = \frac{F_{1i}-\hat{F}_{1i}}{\sigma_{i,(F_{1\_},\hat{F}_{1\_})}}$ so that the density is centered around the prediction. The integral of the standard normal density becomes the standard normal distribution function so that

$$\Phi\left(F_{1i}, \hat{F}_{1i}\right) = \int_{-\infty}^{\left(F_i-\hat{F}_i\right)} g\left(z_i\right) dz_i. \tag{B.3}$$

This can be rewritten as

$$\Phi\left(F_{1i}, \hat{F}_{1i}\right) = \frac{1}{2} erf \left\{ \frac{F_{1i} - \hat{F}_{1i}}{\sqrt{2}\sigma_{i,(F_{1\_},\hat{F}_{1\_})}} \right\}, \tag{B.4}$$

where $erf(\cdot)$ is the Gaussian error function, which can be expanded as

$$\begin{aligned} erf\left(x_i\right) &= \frac{2x_i}{\sqrt{\pi}} {}_1F_1\left(\frac{1}{2}, \frac{3}{2}, -x_i^2\right) \\ &= \frac{2x_i}{\sqrt{\pi}}\left[\sum_{k=0}^{\infty} \frac{(-1)^k x_i^{2k}}{(2k+1)\, k!}\right], \end{aligned} \tag{B.5}$$

${}_1F_1\left(a, b, c\right)$ is a confluent hypergeometric function of the first kind. Plugging (B.5) into (B.4):

$$\Phi\left(F_{1i}, \hat{F}_{1i}\right) = \frac{F_{1i} - \hat{F}_{1i}}{\sigma_{i,(F_{1\_},\hat{F}_{1\_})}} \left[\frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{\left(-\frac{1}{2}\right)^k}{(2k+1)\, k!} \left(\frac{F_{1i} - \hat{F}_{1i}}{\sigma_{i,(F_{1\_},\hat{F}_{1\_})}}\right)^{2k}\right]. \tag{B.6}$$

(B.6) illustrates that the normal distribution is a rescaling of the forecast errors relative to their standard deviation. When $\frac{|F_{1i}-\hat{F}_{1i}|}{\sigma_{i,(F,\hat{F})}} < 1$ then $\frac{|F_{1i}-\hat{F}_{1i}|}{\sigma_{i,(F,\hat{F})}}$ provides a close approximation of (B.6) since the terms inside the brackets collapse to zero for large k. However, when $\frac{|F_{1i}-\hat{F}_{1i}|}{\sigma_{i,(F,\hat{F})}} \geq 1$, then the terms inside the brackets downscale (B.6) so that $\frac{|F_{1i}-\hat{F}_{1i}|}{\sigma_{i,(F,\hat{F})}}$ will be more sensitive to relatively larger errors

B.2 Counterfactual Policy Analysis

Consider a simple representation of damages as:

$$ln\left(p_i y_i\right) = X_i'\beta + \epsilon_i, \tag{B.7}$$

where X_i is a vector of explanatory variables and ϵ_i is assumed to be iid normal. We estimate this model, apply the delta method, and re-scale by prices to get a prediction of real damages

$$\widehat{y}_i \sim \mathrm{IN}\left(y_i, \left(y_i \sigma_{i,\widehat{y}}\right)^2\right). \tag{B.8}$$

The difference between this prediction and some counterfactual, $\widetilde{y}_i$, gives

$$\begin{aligned}
\left(\widetilde{y}_i - \widehat{y}_i\right) &\sim \mathrm{IN}\left(\left(k_i - 1\right) y_i, \mathbb{V}\left(\widetilde{y}_i - \widehat{y}_i\right)\right), \qquad \text{(B.9)}\\
\mathbb{V}\left(\widetilde{y}_i - \widehat{y}_i\right) &= \left(k_i y_i \sigma_{i,\widetilde{y}}\right)^2 + \left(y_i \sigma_{i,\widehat{y}}\right)^2 - 2 * Cov\left(\widetilde{y}_i, \widehat{y}_i\right),
\end{aligned}$$

where $k_i \neq 0$. By the independence and normality assumptions (B.9) can be cumulated as:

$$\sum_{i=1}^{n}\left(\widetilde{y}_i - \widehat{y}_i\right) \sim \mathrm{N}\left(\sum_{i=1}^{n}\left(k_i - 1\right) y_i, \sum_{i=1}^{n}\mathbb{V}\left(\widetilde{y}_i - \widehat{y}_i\right)\right). \tag{B.10}$$

Since we are working with the forecasts, then we typically think of the variance as the residual error variance plus parameter estimation uncertainty as

$$\sigma^2_{i,\widehat{y}} = \left(\beta - \widehat{\beta}\right)' \mathbb{V}\left(X_i\right)\left(\beta - \widehat{\beta}\right) + \sigma^2_{i,\epsilon} \tag{B.11}$$

$$\sigma^2_{i,\tilde{y}} = \left(\beta - \widehat{\beta}\right)' \mathbb{V}\left(\tilde{X}_i\right)\left(\beta - \widehat{\beta}\right) + \sigma^2_{i,\epsilon}. \tag{B.12}$$

Doornik and Hendry (2013) show that this can by approximately estimated as

$$\hat{\sigma}^2_{i,\widehat{y}} = X_i'\left[\widehat{\mathbb{V}\left(\widehat{\beta}\right)}\right] X_i + \hat{\sigma}^2_{i,\epsilon} \tag{B.13}$$

$$\hat{\sigma}^2_{i,\tilde{y}} = \tilde{X}_i'\left[\widehat{\mathbb{V}\left(\widehat{\beta}\right)}\right] \tilde{X}_i + \hat{\sigma}^2_{i,\epsilon}, \tag{B.14}$$

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

$$\widehat{Cov}\left(\widetilde{y}_i, \widehat{y}_i\right) = k_i y_i^2 \left\{\tilde{X}_i'\left[\widehat{\mathbb{V}\left(\widehat{\beta}\right)}\right] X_i + \hat{\sigma}^2_{i,\epsilon}\right\}. \tag{B.15}$$

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

$$\widehat{\mathbb{V}\left(\widetilde{y}_i - \widehat{y}_i\right)} = y_i^2 \left\{\left(k_i\tilde{X}_i - X_i\right)'\left[\widehat{\mathbb{V}\left(\widehat{\beta}\right)}\right]\left(k_i\tilde{X}_i - X_i\right) + \left(k_i - 1\right)^2 \hat{\sigma}^2_{i,\epsilon}\right\}. \tag{B.16}$$

When $\hat{y}_i = \tilde{y}_i$ then $k_i = 1$ and $X_i = \tilde{X}_i$ so (B.16) collapses to zero. Using (B.16) we construct a confidence interval around the difference between the predicted and counterfactual damages.