# EVOLUTION OF SUBJECTIVE HURRICANE RISK PERCEPTIONS: A BAYESIAN APPROACH*

David L. Kelly
Department of Economics
University of Miami
Box 248126 Coral Gables, FL 33124
dkelly@miami.edu

David Letson
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149
dletson@rsmas.miami.edu

Forrest Nelson
Department of Economics
Henry B. Tippie College of Business Administration
University of Iowa
Forrest-nelson@uiowa.edu

David S. Nolan
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149
dnolan@rsmas.miami.edu

Daniel Solís
Rosenstiel School of Marine and Atmospheric Science
University of Miami
4600 Rickenbacker Causeway, Miami, FL 33149
d.solis@miami.edu

May 1, 2011

**Abstract**

How do decision makers weight private and official information sources which are correlated and differ in accuracy and bias? This paper studies how traders update subjective risk perceptions after receiving expert opinions, using a unique data set from a prediction market, the Hurricane Futures Market (HFM). We derive a theoretical Bayesian framework which predicts how traders update the probability of a hurricane making landfall in a certain range of coastline, after receiving correlated track forecast information from official and unofficial sources. Our results suggest that traders behave in a way consistent with Bayesian updating but this behavior is based on the perceived quality of the information received. Official information sources are discounted when a perception of bias and credible alternatives exist.

* Research support grants from the University of Miami Abess Center for Ecosystems Science and Policy, and the University of Miami School of Business Administration are gratefully acknowledged. We would like to thank two anonymous referees, Nuray Akin, Ralphael Boleslavsky, Richard Howarth, Oscar Mitnik, Elizabeth Witham, Nick Vriend, and seminar participants at the American Meteorological Society Meetings, the University of Miami Rosenstiel School of Marine and Atmospheric Science, the Eleventh Occasional Workshop on Environmental and Resource Economics, the Conference on Uncertainty and Learning in the Management of Environmental and Resource Economics at UC-Santa Barbara, and the Wharton Center for Risk Management at the University of Pennsylvania for comments and suggestions. We also thank two anonymous traders for discussing their trading strategies with us in interviews.

## 1. Introduction

Consider a decision maker who solicits advice from several experts, including an official government source, regarding the probability that an adverse event will occur. Such problems are ubiquitous. A stock trader may look at the Securities and Exchange Commission (SEC) and private information sources for balance sheet information to determine the probability of bankruptcy. A bond buyer may consider information from officially-sanctioned rating agencies as well as private sources regarding the probability of default. An individual making nutritional choices may examine government advice (e.g., the USDA food pyramid) and private nutrition books. A regulator may solicit private sector opinions and an internal study regarding the probability of environmental damage. An individual considering evacuation from a hurricane may consult official and private track forecasts to determine the probability that a hurricane will make landfall near the individual. Expert opinions are likely to be correlated, given that each expert observes overlapping data sets. Further, private and official information sources likely have different objectives and incentives. How do decision makers weight private and official information sources which are correlated and differ in accuracy and bias?

Here we use data from a prediction market to study how traders react to official and non-official risk information sources. In particular, we study how traders update beliefs about the probability that a hurricane will make landfall in a certain area ("hurricane risk perceptions") in response to official and non-official hurricane track forecast information.[1] We find that traders were able to spot biases in official information, and used weights consistent with Bayesian

[1] Hurricanes are the most costly natural disaster in the U.S. To mitigate the costs and loss of life of extreme weather, federal agencies have financed weather research programs aiming to improve the accuracy of weather forecasting and to enhance the dissemination of usable weather information (NOAA, 2005).

updating for two information sources. However, traders discounted information from a third source, which was overall the least accurate, but nonetheless provided information that was relatively uncorrelated with the other sources. Information in the third source therefore had a high marginal value.

Because of the difficulty of directly measuring a decision maker's posterior risk perceptions following the solicitation of expert opinion, researchers typically use survey or hedonic methods. Survey research shows that prior perception of risk (Smith, et al. 2001), outside information (Viscusi, 1997; Cameron, 2005), credibility of the source of information (Cameron, 2005 and Viscusi and O'Connor, 1984), socio-economic characteristics (Dominitz and Manski, 1996; Flynn et al., 1994), among other factors, affect risk perceptions. For hurricanes, Baker (1995) uses surveys to study responses to hurricane track forecasts and evacuation notices and finds that updates of official warnings play a major role in shifting stated responses.

Surveys require well-designed monetary payments, or scoring rules, to insure that survey responses accord with actual individual beliefs (for example, Hanson, 2007 contrasts the costs of scoring rules with market-based alternatives). Still, surveys are typically designed with a single information event in mind. Researchers measure respondents' risk perceptions after presenting an information set and associated precision of the information set. Consequently, respondents do not get the opportunity to learn over time which forecasts perform better. Further, the information sources we study are correlated, and so far research using survey methods examines only uncorrelated information. We find that traders give little weight to information that is less accurate but nonetheless valuable since it is relatively uncorrelated with other information sources.

The main alternatives to surveys are hedonic methods, whereby researchers use changes in market prices to reveal changes in risk perceptions. Halstrom and Smith (2005) and Bin and Polasky (2004) are two prominent studies that use changes in housing prices to infer changes in hurricane risk perceptions subsequent to a hurricane event. Such studies must try to control for confounding influences on prices following an adverse event. First, individuals may take on adaptations, such as installing hurricane proof windows, to insulate themselves from future risk. Second, the econometrician may not observe individual heterogeneity in exposure to adverse events or in risk aversion. That is, individuals may be willing to pay different amounts to avoid an adverse event that occurs with a probability that all agree upon. Finally, government regulatory policy may distort prices away from those which represent risk perceptions, by creating a moral hazard problem. This occurs two ways. First, state legislatures may pass regulation which suppresses windstorm insurance premia, *ex ante*. Second, a disaster declaration from the president may provide reimbursement, *ex post*. In both cases, the change in sale prices will not fully reflect the increase in risk perceptions. Some hedonic studies attempt to control for these factors. In particular, Halstrom and Smith (2005) use a near miss hurricane to ensure that rebuilding will not be substantial and find a decrease in housing prices which they attribute to changes in risk perceptions. However, Halstrom and Smith (2005) must still underestimate changes in risk perceptions to some degree, since some homeowners undertake adaptations even if a hurricane is a near miss (especially since the near miss they consider was Hurricane Andrew, a category five hurricane).

The present study proposes an alternative approach to study risk perceptions based on data from a prediction market. In particular, we use the Hurricane Futures Market (HFM) prediction market at the University of Miami. HFM creates securities whose payoffs depend on whether or not a hurricane makes landfall in a specific range of coastline. Traders then trade these securities in an online market.[2] HFM operated during the latter half of the 2005 hurricane season in collaboration with the Iowa Electronic Market (IEM) project at the University of Iowa. The markets ran on the IEM system with traders recruited by HFM. Payoffs are designed so that the price of the security represents the traders' subjective belief of the probability that the event occurs.[3] Hence, the price of the security equals the traders' risk perception.

Prediction markets are well suited to reveal risk perceptions. Because traders win or lose real dollars, they have a strong incentive to reveal, through their trades, their true risk perceptions. Further, by design prediction markets are free of confounding influences from other aspects of risk, such as adaptations or moral hazard. In addition, with hedonic studies, a fall in a sale price may overestimate changes in risk perceptions, because the seller may instead be very risk averse with respect to a small increase in risk. In contrast, losses are small with a prediction market (maximum of $100 in HFM), so approximate risk neutrality is more plausible.

Forty five participants made at least one trade in HFM. Like most prediction markets (and survey populations), HFM traders are not a representative of the general population. In fact, all traders have some interest in hurricanes and/or meteorology (many were undergraduate or graduate students in meteorology). The general population may react differently to new information. Of course, some of the examples given in the first paragraph pertain to experienced decision makers, and some hurricane decision makers (e.g., broadcast meteorologists, traffic engineers) are experienced as well.[4]

[2] See Wolfers and Zitzewitz (2004) for a survey and introduction to prediction markets.
[3] In particular, we are assuming payoffs are small enough so that risk neutrality is a reasonable approximation, that the discount rate is close to one, and security payoffs are not correlated with traders' marginal utility of wealth. See Wolfers and Zitzewitz (2006) for a formal justification.
[4] One disadvantage may be a lack of liquidity in the market, but Tetlock (2007) argues that uninformed traders in thick markets inhibit information revelation.

Lee and Moretti (2009) study the impact of polls on a presidential election prediction market within a Bayesian learning context. Like our paper, they find evidence in favor of Bayesian learning. In particular, they find more precise polls receive more weight by traders. Our results extend upon their study by considering correlated information sources and a (potentially biased) government information source. In addition, although they find that polls, rather than outside information, are the primary drivers of risk perceptions, information revelation is more controlled with hurricanes since new information is revealed only every six hours, when track forecasts are released. Finally, we consider multiple prediction markets and are thus able to control for hurricane-specific fixed effects.

Other studies focus on environmental risk perceptions. Oberholzer and Mitsunari (2006) find that Toxic Release Inventory reports of toxic emissions a moderate distance from a home cause the price to fall, which they attribute to upward adjustments of risk perceptions. A series of papers by Viscusi (Viscusi and O'Connor, 1984; Viscusi and Magat, 1992; Viscusi, 1997; and others) study various environmental risks. For example, Viscusi (1997) studies air pollution and cancer and finds individuals give too much weight to forecasts indicating a high risk of cancer, which they call 'alarmist' learning. Cameron (2005) asks students to forecast future temperatures and studies how risk perceptions change after introducing new information. Students come close to Bayesian learning but place too much weight on priors when forecasts diverge.

Cameron (2005) and Viscusi (1997) consider government and private information sources. Cameron (2005) finds that perception of bias leads to lower weight placed on a particular information source, while Viscusi (1997) finds neither government nor industry sources were more credible. In contrast, we find that traders in our prediction market discount the official National Hurricane Center (NHC) forecast, when it is likely to be biased. In particular, traders discounted the NHC forecast when the NHC forecasted landfall close to an urban area, concluding the NHC was biased to avoid type II (false negative) error.

In general, we find traders behave in a way consistent with Bayesian updating with respect to two forecasts, but essentially ignore a third forecast. The third forecast is the least accurate forecast, and yet provides useful information because it is relatively uncorrelated with the other forecasts.[5] Thus, while we confirm previous results that indicate traders can appropriately weight information sources according to their accuracy, we show this result does not extend to weighting information sources according to their correlation structure, which is more complex.

Overall, traders are remarkably accurate forecasters. Indeed, traders correctly predict a hurricane will or will not make landfall in one of eight Gulf or Atlantic regions with 84% accuracy. The most accurate forecast, the NHC forecast, correctly predicts whether or not a hurricane will make landfall in one of eight regions with 81% accuracy. Traders are more accurate than the NHC for storms more than five days from landfall (69% to 54%), but less accurate for storms two days or less from landfall (90% versus 100%).

Finally, we conduct an *ex-post* test of market efficiency by grouping securities with similar prices and measuring the fraction of securities in the group which eventually pay off (see for example Tetlock, 2004). We find that although price and average payoff are close, a 'favorite-longshot' bias exists in that traders could mildly profit by buying securities with a price near one. The favorite-longshot bias is also evident from the relatively poor performance of the traders for storms close to landfall, which typically have a price very close to one. Jullien and

[5] The third forecast uses statistical information from previous hurricanes, while the other forecasts use mainly physics equations. Thus the third forecast contains different information.

Salanie (2000) and others find the favorite-longshot bias in sports wagering prediction markets, but Tetlock (2004) finds no favorite-longshot bias in financial prediction markets and argues that one possible explanation is that those who bet on sports may be inexperienced with prediction markets. Our results are consistent with this argument in that our traders, while experts in meteorology, have little experience with prediction markets.

The rest of this article is organized as follows. The next section gives an overview of the Bayesian approach in studying risk perception updating analysis, followed by a description of the data and the empirical model. Then, we present and discuss the empirical results. The last Section presents some concluding remarks along with some suggestions for further research.

## 2. Risk Perception Updating Model: The Case of Hurricanes

In the Bayesian framework, new information causes traders to update the probability that a certain hypothesis (a hurricane makes landfall in a certain area in our case) is true.[6] Assume the true probability of hurricane $h$ of type $k$ making first landfall in coastline range $j$ is $P^*$.[7] Note that all of the parameters below and $P^*$ will depend on $j$ and $k$, but we suppress this dependence where no confusion is possible. The true probability is unknown to traders. Since a hurricane of type $k$ will either make landfall in range $j$ or not,[8] traders can view this event as a Bernoulli distributed random variable. That is, each hurricane of type $k$ is a draw from a Bernoulli urn in which $P^*$ is the probability of 'success,' in that the hurricane does make landfall in range $j$. Traders have prior beliefs that $P^* \sim \text{BETA}(\alpha,\beta)$. The beta distribution is particularly advantageous since it allows for a wide variety of density function shapes. The mean of the beta

[6] See Bolstad (2004) for a detailed review of the Bayesian theory.

[7] Hurricane type characteristics may include Atlantic versus Gulf storm, wind speed, and/or the day of the year when the storm formed.

[8] Because storms may straddle more than one range, we define as our trigger event the location where the storm center makes its first U.S. landfall.

distribution is $\alpha/(\alpha+\beta)$, meaning the prior distribution is equivalent to $\alpha$ out of $\alpha+\beta$ draws indicating success. If the prior was formed from previous, similar hurricanes, then the prior indicates $\alpha/(\alpha+\beta)$ fraction of hurricanes of type $k$ ended up making first landfall in range $j$.

Next, suppose traders receive hurricane track forecast information at time $t$. We can view $t$ in six hour increments, since all track forecast are released every six hours. Each track forecast $i$ contains a set of predicted latitude and longitude positions over time. Let $z_{it} = z(\Omega_{it})$ be the traders' belief of the probability of landfall in range $j$ given the latitude and longitude information $\Omega_{it}$ of track forecast $i$ at time $t$.[9] We can view $z_{it}$ as the fraction of $n_{it}$ draws from the Bernoulli urn which indicate success, where $n_{it}=q(\Omega_{it})z_{it}(1- z_{it})$ and $q(\Omega_{it})$ is the precision (or inverse of the variance) of $z_{it}$.[10] Each track forecast is therefore a realization, $z_{it}n_{it}$, of a binomial random variable with parameters $P^*$ and $n_{it}$. The precision varies by track and the time the track forecast was released. Track forecasts vary in their accuracy, and all track forecasts become more accurate at predicting whether a hurricane will make first landfall in a security range as hurricanes approach the coastline. Given our distributional and information assumptions, it is well known (see for example DeGroot, 1970) that, if the track forecasts are independent, the posterior distribution is also beta, with:

$$\alpha_t = \alpha + \sum_{i=1}^{I} z_{it} n_{it} \quad , \quad \beta_t = \beta + N_t - \sum_{i=1}^{I} z_{it} n_{it} \, , \tag{1}$$

$$P_t = \mathrm{E}[P^* \mid H] = \frac{\alpha + \sum_{i=1}^{I} z_{it} n_{it}}{\alpha + \beta + N_t} \, . \tag{2}$$

[9] We specify $z(\Omega_{it})$ precisely in Section 3, but the theoretical model only requires that a function $z$ exists.

[10] We are thus assuming the information content of each track forecast is known (indeed the only uncertain parameter is $P^*$). This assumption is standard in the literature (e.g. Viscusi, 1997), but of course the trader's actual environment is likely considerably more uncertain.

Here $P_t$ is the expectation of $P^*$ conditional on the track forecast information at time $t$, $I$ is the total number of track forecasts and $N_t = \sum_{i=1}^{I} n_{it}$ represents the information contained within the track forecasts. Equation (2) may be decomposed into a linear weighted average of the priors and the information provided by each track forecast, with the weights being equal to the relative information content of each track forecast. Let $D_t = \alpha + \beta + N_t$ be the total precision of the prior and track forecast information. Then:

$$P_t = \frac{\alpha+\beta}{D_t} \cdot \frac{\alpha}{\alpha+\beta} + \frac{n_{1t}}{D_t} z_{1t} + \ldots + \frac{n_{It}}{D_t} z_{It}. \tag{3}$$

Equation (3) shows that the information within each track forecast implies a predicted probability that the hurricane will make first landfall in range $j$, and that the posterior probability is a weighted average of the predicted probabilities. The weights equal the relative information content of each track forecast.

Equation (3) assumes the track forecasts are independent. In fact, the NHC forecast is an expert opinion forecast which explicitly considers other track information. Suppose a known fraction $m_{ijt}$ of the draws tracks $i$ and $j$ made from the urn are common (which draws are common is unknown). We therefore have overlapping information sets (see for example Clemen, 1987 and Zeckhauser, 1971). We can interpret $m_{ijt}$ as a correlation measure, since the correlation between $z_{it}$ and $z_{jt}$ is $m_{ijt} / \sqrt{n_{it}}\sqrt{n_{jt}}$ . Clemen (1987) shows that, when information sources are correlated binomial random variables and the prior is uninformative, the posterior is a mixture of beta distributions. Estimation of mixture distributions is possible (Leroux, 1992), but especially complicated here since each observation has potentially a different mixture of betas. Viscusi (1997) suggests weighting binomial information sources as if the information sources were draws from a correlated multivariate normal distribution. He speculates that the difference between using the weights of the normal distribution and the weights obtained by applying Bayes rule when the prior is beta and the information sources are correlated binomial draws is likely small. We show in appendix 2 that, for values of $N_t$ typically in our data, that in fact the errors are small and therefore adopt Viscusi's suggestion. That is, let $\hat{D}_t = e'V_t^{-1}e$, where $e$ is a unit vector and $V_t$ is the covariance matrix:

$$V_t = \begin{bmatrix} 1/(\alpha+\beta) & 0 & 0 & 0 \\ 0 & 1/n_{1t} & m_{12t}/n_{1t}n_{2t} & m_{13t}/n_{1t}n_{3t} \\ 0 & m_{12t}/n_{1t}n_{2t} & 1/n_{2t} & m_{23t}/n_{2t}n_{3t} \\ 0 & m_{13t}/n_{1t}n_{3t} & m_{23t}/n_{2t}n_{3t} & 1/n_{3t} \end{bmatrix}, \quad e = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

We assume new track information is uncorrelated with the prior. Thus, if information sets are overlapping, and assuming the underlying distribution is normal, the weights are computed via:

$$P_t = \frac{1}{\hat{D}_t} e'V_t^{-1} \cdot \left[\alpha/(\alpha+\beta), z_{1t}, \ldots, z_{It}\right]'.$$

$$P_t = w_{0t} \cdot \alpha/(\alpha+\beta) + w_{1t} \cdot z_{1t} + \ldots + w_{It} \cdot z_{It} \qquad (4)$$

Here $[w_{0t}, \ldots, w_{It}] = e'V_t^{-1}/\hat{D}_t$ are the weights on the prior and track information of security $j$ for hurricane $h$ at time $t$. The conditional expectation remains a linear weighted average of the priors and track forecasts, but the weights account for the probability that information is redundant.

Equation (4) is closely related to Clemen (1987) and Viscusi (1997), who study beta-binomial models. However, all empirical papers assume independent information events. In Section 4 we estimate a version of equation (4) with non-zero correlations, with maximum likelihood, using private and official track forecast data, and prediction market data for the conditional probability $P_t$. In particular, the price of a security which pays \$1 if a hurricane makes landfall in range $j$ trades at price equal to $P_t$, since otherwise a trader could make positive expected profits by buying the security if the price was less than $P_t$, or selling if the price was greater than $P_t$.[11]

## 3. Data

To analyze hurricane risk perceptions, we use data gathered from the Hurricane Futures Market (HFM) project at the University of Miami. When the NHC officially names a tropical storm, HFM creates a market for the storm, which allows traders to buy and sell ten securities whose payoff is conditional on where the storm makes landfall. If the storm is north of a dividing line,[12] the storm is considered in the Atlantic region. Otherwise the storm is in the Gulf region. For an Atlantic region storm, eight securities (labeled A1-A8) pay one dollar if a hurricane makes first landfall within a particular range of U.S. coastline.[13] Securities represent disjoint coastline ranges, and the union of ranges for all securities is the entire U.S. Atlantic coastline from Florida to Maine. Additionally, an 'expires' security, AX, pays one dollar if the storm expires without making U.S. landfall,[14] and a final security, AN, pays one dollar if the storm moves southwest into the Gulf region. For Gulf region storms, eight securities (G1-G8) have coastline ranges which cover the U.S. Gulf coast from Florida to Texas. An expires security, GX, pays if the storm expires at sea or makes landfall in mainland Central America outside the U.S., and a final security, GN, pays if the storm moves North into the Atlantic region. Coastline

[11] Appendix 3 discusses the required conditions for the trade price to equal the trader's conditional probability.

[12] Atlantic storms are those located north and east of the imaginary line extending from Key Largo, Florida (25.25°N, 80.30°W) through the Lesser Antilles (15°N, 65°W) and beyond. Specifically, a storm is designated an Atlantic storm if it forms east of 80.3°W and its latitude satisfies the inequality: latitude > 10.25 · (longitude - 65) / 15.3 + 15.0. Gulf storms are those forming west of 80.3°W or south of that same line when they are named.

[13] NHC provides an exact latitude and longitude corresponding to the site where the center of the hurricane first lands.

[14] A storm is defined to 'expire' if it has not made U.S. landfall or crossed the dividing line and the NHC issues its final advisory for that storm when it is still over the ocean, or over non-U.S. land.

ranges were computed so that since 1949 an approximately equal number of tropical storms or hurricanes made first landfall in each coastline range. Figure 1 is a map of the Eastern U.S. coastline which shows the landfall range of each security and the dividing line.

HFM creates multiple markets if more than one storm is present, and creates more than one market for an individual storm if the storm crosses the dividing line or returns to the ocean after making an initial landfall.

HFM data cover storms for later half of the 2005 season. Many storms and securities elicited little or no trading activity. In such storms the price of a particular security is close or equal to one dollar, and the other securities have a price near zero. Traders put probability near or equal to one on a particular security paying off (typically the "expires" security). Since such storms apparently have no subjective risk, we exclude them from the sample. Our rule is to exclude storms with less than 20 trades. This leaves 4 usable storms, 13 securities, and 445 trades.[15] Thirty two traders made at least one trade in one of the 13 usable securities. Tables 1-3 present summary statistics for the 2005 Atlantic hurricane season and for security prices.

Consider as an example the storm Ophelia, for which a market was created September 7, 2005 and which expired without making landfall on September 16, 2005. Figure 2 presents the evolution of the two securities most likely to pay off: A5 pays one dollar if first landfall occurs in a range of coastline which includes part of North and South Carolina (see Figure 1), and AX pays one dollar if Ophelia expires without making landfall. Initially, A5 traded at a price equal to \$0.10, indicating that traders' subjective risk assessment was that Ophelia would make first landfall in A5 with probability equal to 0.1. New information from track forecasts then arrived, indicating that it was more likely Ophelia would make landfall in A5. Traders then revised their

[15] Our data set is of similar size to other estimates of risk perceptions using survey data (see for example, Viscusi, 1997 or Cameron, 2005).

subjective beliefs upward, eventually to a peak of 0.85 on September 13, as Ophelia neared the Carolina coast. However, Ophelia then turned Northeast and went out to sea, resulting in a decrease in the subjective risk of Ophelia making landfall in A5 to zero by September 15.

Regarding the operational details of HFM, the market consisted of 45 traders. Each trader began with $100 of research funding in their account. No minimum number of trades was required, so traders could make no trades and receive $100. All traders made at least one trade, however. A set of all securities, which pays $1 with probability one, may be purchased at any time from HFM for $1. The securities purchased from HFM form the supply of securities. Traders may also post limit orders to buy or sell a security at a specific price. Traders see the highest buy and lowest sell order, and may accept an offer to buy or sell, creating a trade. HFM records the time, date and price of each trade. Only one trader lost the full $100, and was then out of the market, since traders could not use their own funds. At the end of the season, traders received checks equal to their account value. For more details on HFM, see http://hurricanefutures.miami.edu/.

For the information sources, we collected latitude and longitude data for three standard track forecasts. These track forecasts become available about every 6 hours, being released to the public either shortly before or at the times of midnight, 6 am, noon, 6 pm, Greenwich Mean Time (GMT). All times reported in this paper are GMT. Thus, unlike presidential races or stock markets, information events are easily identified as occurring every six hours. The first track forecast is NOAA's National Hurricane Center forecast, denoted NHC. As few as four NHC hurricane experts consider data from many separate track forecasts (including both of our other track forecasts), and through consensus create the NHC track forecast. Thus, the NHC forecast is a standard expert opinion-type forecast and is also the official government forecast that competes with private forecasts. NHC forecasts have become increasingly accurate over the years, due to computational advances, more data, and improved physical models (Franklin, et. al., 2003). A three-day track forecast today is about as accurate as a two-day track forecast 20 years ago. The mean absolute error for a five-day track forecast is 283.7 nautical miles, which improves to 108.6 nautical miles for a two-day forecast, and 59.6 nautical miles for a one-day forecast (60 nautical miles equals one degree latitude).

The second track forecast is the NOAA/Geophysical Fluid Dynamics Laboratory (GFDL) forecast model (Bender, et. al., 2007). The GFDL model is a structural model (known as a 'dynamic model' in the hurricane literature). Structural models use numerical solutions to physics equations. GFDL forecasts are widely available on the web.

The third track forecast is the Climatology and Persistence (CLP5) forecast model (Aberson, 1998). CLP5 is a purely statistical regression model that forecasts using direction of motion, location, storm intensity, and day of the year information, using parameters estimated from data on previous hurricanes. CLP5 also proxies for basic information about a storm such as the storms current position and heading. CLP5 is widely available in tracking software and on the internet.

Consequently, our data contain three representative models: one expert forecast, one structural model, and one statistical model. Although other models are available (see http://www.nhc.noaa.gov/modelsummary.shtml for details), they are typically either structural, statistical, or a combination of both, and are thus unlikely to add much in the way of information not contained in the models we use. Interviews with the traders revealed that they were using GFDL and were aware of CLP5. Traders seemed to regard CLP5 as too inaccurate to pay much attention to, yet traders did claim to pay attention to the storms' position, heading, speed, and other characteristics upon which CLP5 is based.

Each track forecast contains a set of predicted latitude and longitude positions over time. Each predicted latitude and longitude position is an *l*-hour ahead forecast. Tracks vary in the number of hour ahead forecasts they report, but no forecasts are greater than 126 hours ahead. Table 4 reports forecast accuracy up to 120 hours ahead in 12 hour increments.

As noted in Section 2, traders must convert the point forecasts into probabilities of landfall, with associated precision. Appendix 1 shows how we calculate these probabilities. The probabilities depend closely on the accuracy of the point forecasts and the implied landfall locations. The probability $z$ rises as the predicted landfall location nears the center of range $j$. If the predicted landfall location is in range $j$, then the probability rises with the accuracy of the point forecast. Similarly, precisions vary by track and time since track forecasts vary in their accuracy, and all track forecasts become more accurate as hurricanes approach landfall. Table 5 gives summary statistics for the probability data.

As an example, Figures 3a-c plot the track forecasts for Hurricane Wilma, from October 17-23, 2005. In the graph, the most Northeast marker (square, circle, or plus) is the five-day ahead forecast. From the graphs, on October 17 all three five-day forecasts agreed that Wilma would still be in the Gulf of Mexico. However, the next day GFDL predicted Wilma would land in G7. NHC and CLP5 predicted Wilma would still be at sea, but CLP5 moved closer to the G8 coastline. The implied landfall probability of G7 for GFDL was only 0.24, however. For all tracks, five-day forecasts have large errors. Indeed, the standard deviation of the GFDL five-day forecast error is more than 4.5°, enough to move the landfall to nearly the border between Alabama and Florida. Traders were apparently considerably more confident than the historical accuracy of the GFDL forecast implies, however, since the trade price indicated Wilma would hit G7 with probability 0.40 and G8 with probability 0.45 (GFDL predicted G8 with probability 0.23). The next day, the NHC forecast predicted G8 (specifically, the probability of G8 rose from 0.28 to 0.37) and GFDL predicted the storm would be just off the G8 coastline. The price of G8 rose to 0.7, while the price of G7 dropped to 0.25. Again, traders were considerably more confident than the historical accuracy of the forecasts implied. The price eventually neared one as Wilma neared the G8 coastline, where it eventually made landfall. It is interesting to note that traders appeared overconfident, and yet their forecasts proved correct in this case. Interviews with traders subsequent to the 2005 season indicated traders did not view track forecasts of five days or more ahead as informative, yet their trades indicated surprising confidence. In Section 5 we estimate whether or not traders are systematically overconfident.

Figures 4a-c presents a second example, track forecasts for Hurricane Rita for the dates during which trades occurred (September 21-23, 2005). Rita is interesting because the three September 21, 12 pm forecasts predicted landfall in locations covered by three different securities (the orange lines). The NHC forecast predicted G1 ($z_{NHC,t}=0.49$), GFDL predicted G2 ($z_{GFDL,t}=0.41$), and CLP5 predicted G3 ($z_{CLP5,t}=0.40$). The forecasts predicted the storm was approximately three days from landfall, yet the landfall predictions are relatively close to borders between securities, and so the probabilities are relatively close to one half.

One hour subsequent to the release of these forecasts, the prices were $P_{G1}$=0.4, $P_{G2}$=0.6, and $P_{G3}$=0.03, indicating traders gave the GFDL forecast the highest weight. Traders apparently discounted CLP5, which predicted G3.[16] Indeed, all the forecasts had a higher probability of G3 than the traders. So traders were considerably more confident in G2 than the forecasts implied. Twelve hours later (green lines), the NHC forecast moved to G2 ($z_{NHC,t}=0.55$), GFDL continued

[16] The weights for CLP5 implied from equation (4) are all above 0.29.

to predicted G2 ($z_{GFDL,t}=0.46$), and CLP5 predicted G3 ($z_{CLP5,t}=0.49$). At this point, the price of G1 fell to 0.1, G2 increased to 0.85, and G3 was 0.05. Thus traders placed more weight on the forecasts (GFDL and NHC) that turned out to be correct, because the hurricane made landfall in G2.[17] Further, traders were more confident than the forecasts would suggest, given the forecasts predicted the storm was still more than two days from landfall.

These examples indicate traders can make sophisticated decisions and look at diverse information. They also indicate some possibility of overconfidence. Although these examples are suggestive, a formal statistical model is needed to ascertain exact weights placed on each forecast, and to test whether or not such weights are optimal in a Bayesian sense.

## 4. Empirical Model and Hypotheses

We estimate an empirical version of equation (4) with three track forecasts:

$$P_{hjt} = \beta_0 + \beta_{0h} \cdot s_{ht} + \beta_1 \left( w_{0,hjt} \frac{\alpha}{\alpha + \beta} \right) + \sum_{i=1}^{3} \beta_{i+1} \left( w_{i,hjt} z_{i,hjt} \right). \qquad (5)$$

Here $\beta = [\beta_0, \beta_{0k}, \beta_1, ..., \beta_4]$ is a vector of parameters to be estimated, and $s_{ht}$ is a hurricane-specific dummy.[18] Here $P_{hjt}$ is the conditional probability of hurricane *h* making landfall in range *j*, at the time *t* that a track forecast was released, equal to the price of security *hj* at time *t*.

Equation (5) requires values for the priors $\alpha$ and $\beta$.[19] We considered both the initial CLP5 forecast and an uninformative (improper) prior ($\alpha=\beta=0$). If $\alpha=\beta=0$, then from equation (4):

[17] That the prices in a few cases sum to greater than one most likely occurs due to thinness in the market. In addition, the data are last trade data, and so trades do not occur at exactly the same time.
[18] Adding a dummy for the first day of trading had little effect on the results.
[19] One could also estimate the priors via empirical or hierarchal Bayesian methods (Bernardo and Smith, 2000), rather than assuming a fixed prior, which is the standard Bayesian methodology used here. We consider this

$$w_{0,hjt}\frac{\alpha}{\alpha+\beta}=\frac{\alpha+\beta}{\hat{D}}\cdot\frac{\alpha}{\alpha+\beta}=\frac{\alpha}{\hat{D}}=0. \tag{6}$$

The uninformative prior provides no information (no draws) about $P^*$ and thus the prior receives zero weight. Hence, the $\beta_1$ term drops out of the regression. The remaining weights simplify to $[w_{1t}, w_{2t}, w_{3t}] = e'\widetilde{V}_t^{-1} / \hat{D}_t$, where $\hat{D}_t = e'\widetilde{V}_t^{-1}e$ and $\widetilde{V}_t$ is the matrix $V_t$ excluding the first row and column, and $e$ is now a 3x1 unit vector. The results are virtually identical for both the uninformed and initial CLP5 forecast prior, so we report results using the uninformed prior.

As in equation (4), the data in equation (5) vary by hurricane, security, and over time. Our calculation of the track forecast probabilities in appendix 1 accounts for security- and time-specific information. For example, if a hurricane is forecast to make landfall at the border between securities, the security with the longer coastline will consequently have a higher probability. In addition, probabilities are relatively large if the forecast predicts landfall in a short period of time within the security range. However, it is possible that we have not considered all hurricane-specific information. For example, one track forecast may be more accurate in Gulf versus Atlantic storms, leading to different weights for different storms. Hence we use hurricane-specific fixed effects, which control for unobserved heterogeneity in the hurricane-specific information traders are exploiting that we have not modeled.

Our dependent variable, $P_{hjt}$, lies on the [0,1] interval. However, direct maximum likelihood estimation of equation (5) is generally not feasible since predicted values of $P_{hjt}$ outside the unit interval have beta probability equal to zero. Thus, the likelihood function is not differentiable at 0 and 1, ruling out gradient based likelihood maximization algorithms. For this reason, we follow the literature (see for example, Ferrari and Cribari-Neto, 2004 or Paulino,

---

possibility in appendix 4. Neither Viscusi (1997) nor Lee and Moretti (2009) estimate the priors. Cameron (2005) takes $\alpha$ and $\beta$ directly from survey data.

2001) and use a logit function to transform the conditional mean to the unit interval. By using this approach, our beta distribution model can now be estimated by maximum likelihood. In particular, we rewrite equation (1), defining the beta distribution as a function of the mean and sample size, and then maximize the log of the likelihood function:

$$\begin{aligned}
&\max_{\phi,\psi}\left\{\sum_{h,j,t}\log\left[\mathrm{Beta}(P_{hjt}, g(\phi\cdot x_{hjt}), g(\psi\cdot x_{hjt}))\right]\right\},\\
&x_{hjt} = [1, s_{ht}, w_{1,hjt}z_{1,hjt}, w_{2,hjt}z_{2,hjt}, w_{3,hjt}z_{3,hjt}]',\\
&g(z) = \frac{1}{1+\exp[-z]}.
\end{aligned} \tag{7}$$

However, the transformation makes $P_{hjt} = g(\phi \cdot x_0)$ a non-linear function of the regressors, which is inconsistent with the theoretical model outlined in Section 2. We therefore present a linear approximation of the regression results using first-order Taylor series approximations of the nonlinear density:[20] $P_{hjt} \approx (g(\phi\cdot x_0) - \phi g'(\phi\cdot x_0)) + \phi g'(\phi\cdot x_0)\cdot x = \beta x$, where $x_0$ is the mean of the independent variables.

Equations (4) and (5) imply that traders are Bayesian if $\beta_2 = \ldots = \beta_4 = 1$, so a test for Bayesian updating corresponds to a test of this restriction. If the constant term is positive and significant, then the price exceeds the Bayesian weighted average of forecasts. This signals either that traders have additional information, or that traders are overconfident. If the constant term is positive and significant but traders predictions are less accurate than the prices imply, then traders are overconfident (for example if, when the price is 0.8, hurricanes make landfall in range *j* less than 80% of the time).

[20] The logistic function is relatively linear in the unit interval, so the errors are small. Using a linear approximation is also a standard assumption. Using a linear approximation of the density function is also standard in the literature (see for example, Lee and Moretti, 2009) .

## 5. Results and Discussion

### a. Bayesian Updating Test

Table 6 summarizes the regression results. From Table 6, column 1, the coefficients for GFDL and NHC are highly significant and close to one, the theoretical value consistent with Bayesian updating. The CLP5 coefficient is nearly zero, indicating traders are ignoring CLP5 information, which is consistent with the statements from traders mentioned in Section 3 indicating their belief that CLP5 was too inaccurate to be useful. However, from a Bayesian perspective, traders are underweighting CLP5. Forecast CLP5 is indeed the least accurate, but the low accuracy is somewhat offset by the low correlation CLP5 has with other forecasts. Thus, the information CLP5 does provide has relatively high marginal value. Our results using the entire data set reject strict Bayesian updating. However, the results are consistent with traders who are Bayesians but are unaware of the value of the CLP5 forecast: the hypothesis $\beta_2=\beta_3=\beta_4=1$ is rejected with $\chi^2=92.5$ (p-value=0.00) and the hypothesis $\beta_2=\beta_3=1$ is not rejected with $\chi^2=2.43$ (p-value=0.22). The constant term is positive and significant, indicating either overconfidence or that traders are using other information besides the three track forecasts.

Hurricane Ophelia illustrates how traders ignored CLP5 information. Inspection of Figures 2 and 5a-c reveals that the price of AX closely followed the NHC and CLP5 forecasts throughout much of the trading. For example, on September 13, both NHC and CLP5 predicted the Ophelia make first landfall in range A5, whereas on September 14, both predicted Ophelia would expire at sea. As expected, the price of A5 fell from \$0.85 to \$0.24. However, during September 8-10, the price of AX fell while the CLP5 forecast was moving east (increasing the probability of AX) and the NHC forecast was moving northwest (increasing the probability of A5). Traders apparently put more weight on the NHC forecast during this period, and a trader would have earned more by assigning more weight to CLP5, since AX eventually paid off. Using data only for Ophelia and excluding trades from September 9-10 results in the CLP5 coefficient being positive and not significantly different from one, supporting the idea that traders ignored CLP5 during these dates.

Subsequent to the trading season, we interviewed several traders. They indicated that CLP5 was a better predictor for Ophelia because Ophelia was a slow moving storm and CLP5 forecasts well for slow moving storms. Furthermore, in their opinion, the NHC did not maximize forecast accuracy, because it faces different penalties for type I (false positive) and type II (false negative) errors. In their opinion, the NHC predicts landfall too often and predicts storms will land near or on an urban center too often.[21] For Ophelia, traders we interviewed felt the NHC was predicting landfall because it feared the consequences of predicting that Ophelia would go out to sea, only to see it make landfall in the Carolinas.

To test this idea, in Table 6, column 3, we interact the Ophelia dummy with the NHC forecast. The coefficient is not significant, indicating traders did not give the NHC forecast significantly less weight for Ophelia. Overall then, even though some traders felt the NHC forecast was biased for Ophelia and CLP5 was predicting AX, traders went with the NHC forecast (especially during September 8-10) because they did not view the CLP5 as providing valuable information.

Wilma provides another test case, this time between GFDL and the NHC. From Figure 3, the NHC consistently forecasted G8, whereas GFDL forecasted G7 on October 17-18, but switched to G8 on October 19, and then trended south towards GN.[22] The NHC forecast was

[21] Powell and Aberson (2001) examine NHC forecasts between 1976 and 2000 and find a bias to avoid type II errors, which they call a "least regret" forecast. Note that while it may be entirely appropriate for the NHC to bias forecasts in this way, our interest is in how traders react to a potentially biased forecast.
[22] Traders discounted CLP5, which predicted mainly G5 and G6 until the very end.

very close to Tampa, a large urban center, whereas GFDL trended south to a less populated area. Prices appeared to closely follow GFDL. The price of G7 declined from $0.44 to trade in the range of $0.10 to $0.25 after October 18, before declining to near zero as GFDL trended toward GN. Similarly, the price of GN increased briefly to $0.35 at the end of the day on October 19 as GFDL began to drift south of Tampa. From October 20-23, the probability of GN for GFDL declined because the effect of the standard error of the forecast narrowing as the storm approached the coast outweighed the effect of the forecast nearing the GN border. The price of GN also declined during the period from October 20-23. The prices are therefore consistent with traders' favoring GFDL over NHC. Interviews with the traders indicated they discounted the NHC forecast because the felt the NHC was compelled to predict landfall near Tampa.[23]

In Table 6, column 4, we formally test this idea by including a term which interacts the Wilma dummy with the NHC forecast. The coefficient is negative and significant as expected, indicating that traders discounted the NHC forecast in favor of GFDL for Wilma.[24] Overall then, the results for Ophelia and Wilma indicate that traders discount the official forecast when they perceive bias and when they perceive the alternative information source is credible.

Turning next to Katrina, although CLP5 and GFDL briefly turned towards G4 very early on (Figures 6a-c), all three forecasts consistently predicted G3, the eventual winner. Traders also favored G3, whose probability never fell below 0.73. Nonetheless, traders seemed to favor G4, assigning G4 probabilities as high as 0.7 during trading,[25] despite the fact that no forecast had the probability of G4 above 0.07 during the period of trading. Katrina made landfall in G3, but

[23] Regional weather patterns indicated there was almost no chance of a landfall near Tampa. Thus traders could effectively rule out model uncertainty as a reason for divergence of the forecasts.

[24] We can reject the hypothesis that the sum of the NHC coefficients equals one at the five percent level of significance.

[25] The sum of the last trade prices was greater than one for about one day. This may reflect illiquidity in the market. In addition, Katrina was the first hurricane with active trading. Therefore, there was probably quite a bit of learning during Katrina trading.

extremely close to the border of G4. Close enough, in fact, that it took a couple of days to determine the winning security. The true probability of G4 is of course unobserved, but most likely greater than the probability indicated by the track forecasts. For consistency, in Table 6, column 5, we included a term which interacted the NHC forecast with the Katrina dummy. As expected the coefficient was not significant, since the forecasts were all in agreement that Katrina would make landfall near an urban area.

Turning next to Rita, from Figures 4a-c, GFDL and NHC generally predicted G2, whereas CLP5 generally predicted G3. Prices appeared to closely track GFDL and NHC, which was correct *ex post* since Rita eventually made landfall in G2. In Table 6, column 6, we included a term which interacted the NHC forecast with the Rita dummy. As expected, the term was not significant. GFDL was generally closer than the NHC to the urban center Houston, so no perception of bias existed.

Overall then, our results indicate some support for Bayesian updating with respect to GFDL and NHC, but with CLP5 being generally underweighted. Furthermore, traders perceived the NHC forecast was biased in cases where the NHC predicted landfall near an urban center and when an alternative information source perceived to be credible predicted landfall elsewhere.

Government information dissemination faces difficult tradeoffs in its principal-agent problem. If the information is unbiased, then the government agency may face a high penalty for not predicting an adverse event that occurs, while if the agency submits biased information it may be ignored. Here, the NHC apparently leans toward releasing biased information to ensure that type II errors will not occur (Powell and Aberson, 2001). Our results indicate that the NHC bias crept into the price of Ophelia AX, even though some traders were aware of it. However, the NHC bias did not affect Wilma security prices, as traders discounted the NHC forecast in favor of GFDL.

### b. Accuracy

Consider as a measure of accuracy the fraction of trades for which $P_{hjt} \geq (<) 0.5$ and the hurricane made (did not make) landfall in range $j$. Table 7 indicates that, by this measure, traders forecast with an 84% success rate. When a hurricane is three days or less from landfall, the percentage rises to better than 90%. Traders are remarkably accurate in their forecasts. For forecasts, we similarly measure the fraction of forecasts for which $z_{hjt} \geq (<) 0.5$ and the hurricane made (did not make) landfall in range $j$. Table 7 indicates that, as expected, the official NHC forecast is the most accurate forecast, whereas CLP5 is the least accurate. Traders are more accurate than the NHC for storms greater than three days from landfall, whereas the NHC is more accurate for storms less than or equal to two days from landfall. Overall traders are slightly more accurate than the NHC.

One possible reason why traders are less accurate for storms near landfall is a 'favorite-longshot bias' (see for example Tetlock, 2004). The favorite-longshot bias occurs when expected returns from betting increase with the probability of winning. Traders could mildly profit by buying securities for which the hurricane is near landfall and forecasted to make landfall in the security range. Such securities have a high price and are thus 'favorites.' Additional evidence of a favorite-longshot bias is presented below.

In Table 7, each trade counts as one observation of trader beliefs. That is, when a buyer and seller agree on a trade price, we take the trade price as a proxy for the average beliefs of the two traders. However, new information is released only every six hours, so trades in the same six hour window at very different prices might reflect illiquidity in the market, rather than new information. For example, 11 trades for Rita, security G3, occurred within three hours of 12 noon on September 23. The average price was \$0.32, while the forecasts ranged from 0.17 to 0.20. However, one trade was at \$0.60. In a more liquid market with less price dispersion, trades would be closer to the mean trade, as the buyer would be able to find a seller for a price less than \$0.60.

In Table 8, we group trades by the nearest forecast release. In particular, since forecasts are released every six hours, each trade is matched to a set of forecasts no more than plus or minus three hours from the trade. We then average all prices that are matched to the same set of forecasts. If a six hour period has no trades, we have no observation for that time interval. Table 8 reveals that the NHC forecast accuracy falls slightly, while the HFM forecast accuracy improves considerably to 94%. Averaging the trades reduces the impact of some less accurate trades that probably would not occur in a more liquid market. HFM still outperforms all three track forecasts. Indeed, we computed HFM forecasts a number of ways and HFM outperformed all track forecasts with the exception of storms very close to landfall. The primary advantage of HFM is in Ophelia and Wilma, when the storms were more than five days from landfall. Thus, as noted in Section 5a, traders are more accurate in situations where the NHC faces a large penalty for type II error.

Still, Tables 7 and 8 do not provide a test of market efficiency. For example, it may be that 94% of storms of type $k$ make first landfall in range $j$, but traders only assess a probability of 0.7. In that case Table 7 would indicate 94% accuracy, but traders would be consistently underestimating the probability of success, and traders could make positive expected profit of \$0.24 by buying the security for \$0.70, with expected payout of \$0.94. Market efficiency implies the current price is the conditional probability, based on current public information, that the hurricane will make landfall in range *j*. Unfortunately, the true conditional probability is unobserved. Instead, we observe only whether or not the security paid off. However, by grouping securities with similar prices, we can estimate the true conditional probability with the fraction of securities that eventually pay off. Therefore we grouped nearby trade prices into 20 equal sized bins,[26] and then for each bin compare the midpoint of the range of prices in the bin with the percentage of actual successes for all trades in the bin's price range. Market efficiency implies the relationship between price and the percentage of actual successes is the 45 degree line.[27]

Figure 7 shows that most observations are near the 45 degree line, but the slope is greater than one. Traders could mildly profit by betting on storms with a high price and selling securities with a low price (buying favorites, selling long shots).[28,29] Hence Figure 7 is consistent with a favorite-longshot bias. These results must be interpreted with caution because of the difficulty of estimating an event with a probability near one without a large data set. Even grouping all trade prices greater than or equal to 0.8 to increase the sample size, however, gives a slope greater than one.

## 6. Concluding Remarks

This study is the first to use prediction market data to study hurricane risk perceptions. Our regression model estimates how individuals update their subjective risk perceptions in response

[26] The bins are of equal size ($0.01-$0.05, $0.06-$0.10, etc) and the results are not very sensitive to the number of bins used.

[27] Assuming a discount factor of one, risk neutrality, that payoffs are uncorrelated with wealth, and no transactions costs.

[28] Interestingly, one trader we interviewed noticed the bias and made significant profits selling securities with a low price. These trades apparently did not completely eliminate the bias, however.

[29] Jullien and Salanie (2000) and others find evidence of a favorite-longshot bias in horse racing, but Tetlock (2004) finds a reverse longshot bias for the case of sports prediction markets and no bias for financial prediction markets.

to private and official sources of hurricane track forecast information. An important issue that we address is how much weight individuals place on competing information sources, as well as their own prior beliefs, as they update their subjective beliefs about hurricane landfalls. We find traders behave in a manner consistent with Bayesian updating with respect to the official (NHC) forecast and a structural forecast model (GFDL), but underweight a statistical model (CLP5).

CLP5 is the least accurate forecast, but receives significant weight in the Bayesian forecast because it is relatively uncorrelated with the other forecasts. Since the value of CLP5 is subtle, it is perhaps not surprising that boundedly rational traders were unable to see the value of CLP5 information. Nonetheless, our results indicate differences between uncorrelated and (until now not examined in the literature) correlated information sources, since the value of correlated information sources is more difficult to ascertain.

Traders display remarkable skill. Traders correctly predict whether a hurricane will or will not make landfall in one of 8 regions for 84% of their trades. If the hurricane is 3 days or less from landfall, the percentage rises to over 90%. When comparing average forecasts made by traders with track forecasts on the same hurricane at roughly the same time, the traders forecast with 94% accuracy compared to 77% accuracy of the best available track forecast (NHC).

Nonetheless, the NHC forecast outperformed traders for storms less than or equal to three days from landfall. Track forecasts are highly accurate when storms are near landfall. Hence, security prices should be near one if the storm is projected to make landfall in the security range and near zero otherwise. But security prices tended to be too low when the landfall probability was near one and too high when the landfall probability was near zero. This behavior is consistent with a favorite-longshot bias.

With regard to official versus private information, traders believed the NHC forecast was biased to avoid type II errors. For Wilma, traders discounted the NHC forecast in favor of GFDL (which turned out to be correct), but for Ophelia traders did not discount the NHC forecast in favor of CLP5 (which turned out to be correct). Traders perceived bias in both cases, but were only willing to discount the NHC forecast when the alternative forecast was perceived as credible.

Several caveats are in order. First, HFM is a thin market. Due to the lack of trades, we cannot introduce other track forecasts traders may be watching, including official forecasts with different time lags. Nonetheless, it is unclear if adding additional noise traders would help or hinder information revelation. Second, our traders are mostly meteorologists, so it is unclear if the results generalize to the general population. Still, a wide variety of experienced decision makers consult possibly correlated and competing official and unofficial information sources.

Our results indicate official information agencies such as the NHC face a difficult tradeoff. Penalties for type II error may lead information agencies to bias information dissemination, but biased official information may be discounted, at least by experienced decision makers, in the face of credible alternatives. It would be interesting to see if this tradeoff extends to other information agencies. We leave this question to future research.

## 7. Appendix 1: Computing the probabilities and precisions

Every six hours, institutions release track forecasts, which give point forecasts of the storm's position at various times in the future. Thus the information in track forecast $i$ released at time $t$, $\Omega_{it}$, consists of a set of $L_i$ pairs of position coordinates, so $\Omega_{it} = \{Lat_{itl}, Lon_{itl}\}$, $l=1 \ldots L_i$. Tracks vary in the number of point forecasts issued at each release. Each trader must at least implicitly convert the information $\Omega_{it}$ into a probability of first landfall in range $j$, $z_{ijt}$, upon which the price of the security is based. Here we explain how we compute the probability, based on the point forecasts and historical mean forecast errors.

The first step is to compute the standard deviation of the point forecast errors. NHC (2008) gives historical mean prediction errors by track and time to landfall. Hurricane track forecasts have become increasingly accurate, thus we consider only the 2000-06 mean absolute errors (2002-06 for CLP5). It is straightforward to show (proof available on request) that, if the latitude and longitude point forecast errors are normally distributed with mean zero and if the variance of the latitude error equals the variance of the longitude error, then the standard deviation in each direction, $\sigma_{il}$, is related to the mean absolute Euclidean distance error, $MAE$, according to: $\sigma_{il} \sim \sqrt{2/\pi} \cdot MAE_{il}$. For example, the third point on the black line of Figure 6c is the 36 hour ahead forecast of the August 28, 12 pm CLP5 Katrina forecast. Thirty six hour ahead CLP5 forecasts have a mean absolute distance error of 315.4 km, which corresponds to a standard deviation of the longitude error of 251.7 km.

We next compute the probability of landfall in range $j$ via a Monte Carlo procedure. The simulated actual longitude position, $\tilde{c}_{itl}$, is a normally distributed random variable with mean equal to the point forecast $Lon_{itl}$ and standard deviation $\sigma_{il}$. However, errors are positively serially correlated over point forecasts. That is, if the hurricane is west of the one day ahead forecast, the hurricane is likely to be west of the two day ahead forecast as well. We model the serial correlation as:

$$\tilde{c}_{it1} = \mathrm{N}(Lon_{it1}, \sigma_{il}^2)$$
$$\tilde{c}_{itl} = \tilde{c}_{it,l-1} + \mathrm{N}(Lon_{itl} - Lon_{it,l-1}, \sigma_{il}^2 - \sigma_{i-1,l}^2) \quad i = 2 \ldots L_i$$

Hence, $\tilde{c}_{itl} - Lon_{itl}$ is mean zero with standard deviation $\sigma_{il}$, but is positively serially correlated over $i$. A simulated actual latitude is computed in an identical manner.

Next, for each simulated actual track, $\tilde{c}_{it} = \{\tilde{c}_{it1}, \ldots, \tilde{c}_{itL}\}$ and the same for latitude, we compute the first landfall, by interpolating between the simulated actual positions, accounting for the curvature of the earth. The first landfall coordinate is the first intersection between the coastline (including the dividing line) and the line formed by interpolating the simulated actual positions. If all forecasts are at sea (over land), the track forecast is extrapolated forward (backward) using the last (first) two forecasts. If all simulated actual positions and the forward extrapolation are at sea, the simulated track is said to predict the storm will expire at sea. The probability is thus equal to the fraction of a large number (1000) of simulated actual tracks which make first landfall in range $j$.

To compute the precision of the probability, $q(\Omega_{it})$, we use a bootstrap procedure. For each track forecast, we have 1000 simulations which either made landfall in range $j$ or did not. From this set, we draw a large number (1000) of samples with replacement of 1000 simulations each and compute the fraction which make landfall in range $j$. We then have a set of 1000 probabilities. The precision is the inverse of the variance of the set of probabilities.

Our methodology requires some fairly sophisticated computations. However, traders are at least aware of the various track forecasts, and how their error varies across tracks and as the hurricane approaches landfall. More straightforward methods of computing probabilities from track forecast information should produce similar regression results.

## 8. Appendix 2: Normal approximation

Here we show that the approximation of the posterior distribution of $P^*$ used in the paper is a reasonable approximation of the actual distribution. The approximation is equal to Beta($P_t$,$N_t$-$P_t$), where $P_t$ is as in equation (4), while Clemen (1987) shows the actual distribution is a mixture of beta distributions. Intuitively the normal approximation works for two reasons. First, all information sources (track forecasts) provide unbiased estimates of the true probability. Therefore, the track forecasts tend to converge on the true value for large $N$ and the weights become irrelevant.

Second, Clemen (1987) shows the mixture of distributions arises from the decision maker (trader) inferring the total successes for all track forecasts by the reports of each track forecast. The total successes, $s$, is unknown since for example if two track forecasts have $m_{12}$=1 and report one success in four draws ($z_i$=1/4), then either both saw one shared draw ($s$=1) or each saw one successful private draw ($s$=2). In contrast, the approximation does not infer the total successes, but instead just constructs a weighted average of the track forecasts. Therefore, the variance of the estimate of $P^*$ is higher. However, with large $N$ the posterior is a mixture of many beta distributions, the weighted average of which tends to be close to normal.

Figure 8 gives an average of 50,000 posterior distributions each of which draws random $z_i$'s from a binomial distribution where the true value is 0.5, and $n_1$=25, $n_2$=30, $n_3$=20, and $m_{12}$=15, $m_{13}$=0, $m_{23}$=0, and $m_{123}$=5. The solid line is a theoretical best case posterior supposing the trader knew the total successes, which has the smallest variance. The mixture of betas is unbiased, but has a higher variance since the trader estimates the total number of successes. The variance of the normal approximation is 1.5% higher still, since the normal distribution does not try to infer the total successes. Figure 9 shows how the approximation error decreases with the total number of draws $n=n_1+n_2+n_3+2m_{123}-m_{12}-m_{13}-m_{23}$. For $n>15$, still well below $n$ for even five day track forecasts, the variance is less than two percent higher than the variance of the mixture of betas. Thus the approximation is reasonably accurate.

## 9. Appendix 3: Correlation between security payoff and marginal utility of wealth

A number of papers give sufficient conditions for prediction market prices to be equal to the traders' subjective probability that the event occurs. Clearly risk neutrality is sufficient (Wolfers and Zitzewitz, 2004). Wolfers and Zitzewitz (2006) show that if the distribution of beliefs is symmetric and demand is symmetric around the point where price equals probability, then prices equal probabilities even if traders are risk averse. These papers have a number of other implicit assumptions, such as no transaction costs and a discount factor equal to one. These assumptions are relatively innocuous in our case: the small stakes make risk neutrality a reasonable assumption, and the security pays off within a week or so of trading, so the discount factor is near one.

However, one assumption is potentially relevant to our application. Wolfers and Zitzewitz (2004) note that, if traders are risk averse and payoffs are correlated with agents' marginal utility of wealth, then prices and probabilities can differ. For example, risk averse traders with at least partially uninsured assets, such as housing, along the coastline of G3 may purchase G3 at a price above the subjective probability of landfall for insurance reasons. If the hurricane makes landfall in G3, the trader's asset losses are offset from the payoff from G3.

Similarly, if traders have assets under risk along the coastline, then GX and AX are positively correlated with traders' wealth. Thus GX and AX would trade below the subjective probability of expiring at sea. Risk averse traders require a premium of a positive expected return to hold the added risk of GX and AX (the principle is the same as the capital asset pricing model, CAPM).

To see this precisely, let $q(k)$ be trader $k$'s subjective probability of GX or AX, $y_x(k)$ be the wealth of trader $k$ if the hurricane expires at sea, and $y_0(k)=\delta y_x(k)$ be the expected wealth of trader $k$ if the hurricane makes landfall. Following the logic of equation (5) in Wolfers and Zitzewitz (2006) we have for log utility:

$$P_t = \int q(k)\frac{y_0(k)}{\bar{y}}F(k)dk$$
$$\bar{y} = \int \big(q(k)y_0(k) + (1-q(k))y_x(k)\big)F(k)dk$$

That is, traders with more wealth at risk than the ex ante average expected wealth demand relatively less of a security whose payoff is positively correlated with their wealth, decreasing the price. We assume no correlation exists between traders' wealth at risk and subjective beliefs.[30] Hence, rewriting gives:

$$P_t = \frac{\delta}{1-(1-\delta)\,\mathrm{E}[q]}\mathrm{E}[q]$$

Thus $P_t<\mathrm{E}[q]$, since the probability for each trader is less than one.

[30] It is possible that traders with more wealth at risk also have an incentive to acquire more accurate priors, but unlikely that traders with more wealth at risk systematically believe hurricanes are less likely to make landfall. In fact, though, $P_t<\mathrm{E}[q]$ as long as the covariance satisfies:

$$\mathrm{cov}(q, y_x) < \frac{(1-\delta)\,\mathrm{E}[q](1-\mathrm{E}[q])\,\mathrm{E}[y_x]}{\delta + (1-\delta)\,\mathrm{E}[q]}$$

We therefore test $P_t$<E$[q]$ for the security AX,[31] by including a dummy variable equal to one if the trade was an AX security. The null hypothesis is that the coefficient on the dummy is negative. The price should be lower for securities positively correlated with assets at risk.

As shown in Table 6, we can reject the hypothesis that the coefficient is negative at the five percent level. We therefore reject that traders are using HFM to hedge wealth at risk from hurricanes.

**10. Appendix 4: Estimation of the priors.**

Our empirical results assume priors are uninformative, or $\alpha=\beta=0$. Another possibility is to estimate the priors which best fit the data. Estimation of priors is known as empirical Bayesian methods. Another alternative we do not consider is hierarchical Bayesian methods, in which the econometrician has priors over $\alpha$ and $\beta$.

To estimate the priors, we take $\alpha$ and $\beta$ in equation (7) as parameters to be estimated, rather than assuming $\alpha=\beta=0$. Column 8 of Table 6 details the results. The estimate of $\alpha+\beta$=1.05 is close to one. The estimated precision of the prior is then equal to about one draw. Thus the econometric results show that the traders priors are not very informative, providing some justification for our assumption of an uninformative prior. The estimate of $\alpha/(\alpha+\beta)$=0.68 indicates the prior amounts to a very noisy prediction that the hurricane makes landfall in the coastline range 68% of the time. The $\chi^2$ value drops with the addition of $\alpha$ and $\beta$, indicating that estimating the priors has not significantly improved the fit of the model.

As for the weights on the track forecasts, the weight on GFDL increases slightly and the weight on NHC decreases slightly and the standard error of both weights decreases, so we can now reject the joint hypothesis that both weights are equal to one. However, the results remain

[31] We have no trades for the GX security.

qualitatively unchanged in that traders still put essentially no weight on CLP5, and that the GFDL and NHC weights remain close to one.

## 11. Appendix 5: Tables and Figures

**Table 1**: Summary statistics for the 2005 Atlantic hurricane season.

| | |
|---|---|
| Number of Tropical and Subtropical Storms | 28 |
| Number of Hurricanes | 15 |
| Number of Major Hurricanes (Cat. 3-5) | 7 |

**Table 2.** HFM Trade Data for year 2005: Summary Statistics. Hurricanes can potentially have multiple landfalls and thus multiple markets. Forty five traders participated, of which 32 made at least one trade in the subset of securities with at least 20 trades. All traders began with $100.

| Summary Statistic | Number | Summary Statistic | mean | Std. Dev. | Max. | Min. |
|---|---|---|---|---|---|---|
| Storms with markets | 11 | Contracts | 730.6 | 1,059.9 | 4,150.0 | 9.0 |
| Securities with >20 trades | 13 | Ending Balance ($) | 103.0 | 52.6 | 207.6 | 0.0 |
| Storms with > 20 trades | 5 | | | | | |

**Table 3.** Summary statistics for transaction prices by storm and security.

| Storm (Intensity) | Trades | Mean | Std. Dev. | Maximum | Minimum |
|---|---|---|---|---|---|
| Katrina-Atlantic (Cat 1) | | | | | |
| Security A2 | 20 | 0.456 | 0.215 | 0.700 | 0.020 |
| Katrina-Gulf (Cat 5) | | | | | |
| Security G3 | 22 | 0.894 | 0.063 | 0.980 | 0.730 |
| Security G4 | 39 | 0.400 | 0.161 | 0.700 | 0.001 |
| Ophelia (Cat 1) | | | | | |
| Security A5 | 29 | 0.235 | 0.218 | 0.750 | 0.001 |
| Security AX | 46 | 0.420 | 0.198 | 0.895 | 0.050 |
| Rita (Cat 5) | | | | | |
| Security G1 | 36 | 0.379 | 0.162 | 0.700 | 0.015 |
| Security G2 | 73 | 0.766 | 0.150 | 0.990 | 0.150 |
| Security G3 | 63 | 0.168 | 0.145 | 0.600 | 0.001 |
| Wilma (Cat 5) | | | | | |
| Security G7 | 26 | 0.247 | 0.148 | 0.440 | 0.010 |
| Security G8 | 27 | 0.736 | 0.182 | 0.990 | 0.350 |
| Security GN | 32 | 0.176 | 0.079 | 0.350 | 0.010 |

**Table 4.** Standard deviation of longitude and latitude distance error in km by track and hours ahead. Source: authors' calculations from data published by NHC (2008).

| Track | Hours ahead | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 12 | 24 | 36 | 48 | 72 | 96 | 120 |
| GFDL | 64.6 | 112.1 | 177.1 | 257.7 | 353.6 | 486.9 | 552.1 | 896.3 |
| NHC | 17.8 | 89.6 | 156.1 | 219.0 | 284.4 | 422.2 | 558.9 | 743.1 |
| CLP5 | 31.3 | 131.2 | 270.8 | 446.0 | 596.9 | 872.7 | 1088 | 1297 |

**Table 5.** Summary statistics for track forecasts. The top number in each cell is the mean and the bottom number is the standard deviation. For columns 2-4, each row considers all 12 hour ahead track forecasts occurring between the start and end of trading for the security given in the first column. Columns 2-4 give the average distance in km to the closest point in the range of coastline given by the security listed in the rows. If the 12 hour ahead forecast is over land, the distance is set to zero. For AX, the distance is the average minimum distance to the US coastline. For track forecasts that eventually cross the coastline, columns 8-10 are computed by interpolating between the hour ahead forecast that is last over water and the first hour ahead track over land. If the 120 hour ahead forecast does not cross land, the track forecast is interpolated forward. If the forecast predicts the storm will expire at sea, the hour ahead forecast closest to landfall is used.

| Storm | Distance to security range (km) | | | Probability of landfall in security range | | | Predicted hours to landfall | | |
|---|---|---|---|---|---|---|---|---|---|
| | GFDL | NHC | CLP5 | GFDL | NHC | CLP5 | GFDL | NHC | CLP5 |
| **Katrina-Atlantic** | | | | | | | | | |
| Security A2 | 136.8 | 126.3 | 119.1 | 0.10 | 0.22 | 0.27 | 33.0 | 26.5 | 42.0 |
| | 114.1 | 123.4 | 136.9 | 0.02 | 0.09 | 0.17 | 19.6 | 16.8 | 38.5 |
| **Katrina-Gulf** | | | | | | | | | |
| Security G3 | 119.3 | 127.5 | 150.9 | 0.74 | 0.80 | 0.55 | 34.0 | 36.0 | 52.0 |
| | 122.2 | 130.1 | 142.5 | 0.13 | 0.09 | 0.20 | 11.8 | 10.7 | 19.6 |
| Security G4 | 73.3 | 75.5 | 109.8 | 0.12 | 0.11 | 0.05 | 34.0 | 36.0 | 52.0 |
| | 60.9 | 72.8 | 74.3 | 0.05 | 0.04 | 0.02 | 11.8 | 10.7 | 19.6 |
| **Ophelia** | | | | | | | | | |
| Security A5 | 117.0 | 126.1 | 142.8 | 0.22 | 0.23 | 0.18 | 54.0 | 52.8 | 43.4 |
| | 110.8 | 120.6 | 130.2 | 0.16 | 0.18 | 0.18 | 32.7 | 39.3 | 27.2 |
| Security AX | 174.9 | 185.2 | 192.7 | 0.27 | 0.37 | 0.42 | 54.0 | 52.8 | 43.4 |
| | 99.5 | 105.6 | 100.2 | 0.20 | 0.24 | 0.23 | 32.7 | 39.3 | 27.2 |
| **Rita** | | | | | | | | | |
| Security G1 | 617.7 | 610.9 | 634.0 | 0.18 | 0.19 | 0.10 | 45.2 | 43.4 | 49.8 |
| | 339.2 | 331.9 | 343.8 | 0.14 | 0.19 | 0.08 | 22.4 | 20.5 | 20.1 |
| Security G2 | 365.1 | 359.8 | 383.1 | 0.56 | 0.62 | 0.30 | 45.2 | 43.4 | 49.8 |
| | 300.0 | 290.0 | 301.1 | 0.15 | 0.19 | 0.25 | 22.4 | 20.5 | 20.1 |
| Security G3 | 79.7 | 72.0 | 84.8 | 0.23 | 0.18 | 0.40 | 45.2 | 43.4 | 49.8 |
| | 127.9 | 124.9 | 135.3 | 0.08 | 0.08 | 0.18 | 22.4 | 20.5 | 20.1 |
| **Wilma** | | | | | | | | | |
| Security G7 | 307.5 | 291.5 | 322.7 | 0.18 | 0.18 | 0.08 | 85.5 | 77.1 | 109.3 |
| | 136.6 | 131.3 | 144.4 | 0.10 | 0.08 | 0.06 | 36.9 | 34.5 | 30.9 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Security G8 | 332.8 | 317.2 | 343.1 | 0.38 | 0.46 | 0.14 | 85.5 | 77.1 | 109.3 |
| | 183.1 | 170.1 | 187.3 | 0.24 | 0.25 | 0.17 | 36.9 | 34.5 | 30.9 |
| Security GN | 492.0 | 468.2 | 498.5 | 0.22 | 0.21 | 0.15 | 85.5 | 77.1 | 109.3 |
| | 243.3 | 240.7 | 250.1 | 0.10 | 0.11 | 0.09 | 36.9 | 34.5 | 30.9 |

**Table 6.** Maximum likelihood estimation, all storms. Column 2 is random effects, all other columns include hurricane-specific fixed effects (implemented using dummy variables). Columns 3-6 include interaction terms, column 7 includes a dummy variable for AX trades, and column 8 estimates the prior parameters using empirical Bayesian methods. Independent variables equal $z_{it}$, *i= CLP5, GFDL, NHC.* $^{**}$,$^{***}$ indicates significance at the 5% and 1% level, respectively, and standard errors are in parenthesis. Except for $\alpha$ and $\beta$, we report first-order Taylor series approximations of the mean of the density. All regressions have 417 observations.

**Econometric Specification**

| **Coefficient** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|---|---|---|---|---|---|---|---|---|
| **Constant** | $0.27^{***}$ (0.02) | $0.27^{***}$ (0.02) | $0.26^{***}$ (0.05) | $0.27^{***}$ (0.05) | $0.27^{***}$ (0.06) | $0.27^{***}$ (0.05) | $0.28^{***}$ (0.03) | $0.25^{**}$ (0.11) |
| **GFDL** | $1.13^{***}$ (0.13) | $1.13^{***}$ (0.13) | $1.17^{***}$ (0.13) | $1.12^{***}$ (0.12) | $1.14^{***}$ (0.13) | $1.06^{***}$ (0.13) | $1.11^{***}$ (0.13) | $1.19^{***}$ (0.07) |
| **NHC** | $0.89^{***}$ (0.11) | $0.90^{***}$ (0.10) | $0.95^{***}$ (0.11) | $0.81^{***}$ (0.11) | $0.92^{***}$ (0.11) | $1.05^{***}$ (0.14) | $0.88^{***}$ (0.11) | $0.80^{***}$ (0.06) |
| **CLP5** | 0.00 (0.12) | 0.00 (0.12) | 0.00 (0.12) | -0.17 (0.13) | -0.02 (0.12) | -0.05 (0.12) | 0.00 (0.12) | -0.16 (0.12) |
| **OPHxNHC** | | | -0.33 (0.27) | | | | | |
| **WILxNHC** | | | | $-0.57^{***}$ (0.19) | | | | |
| **KATxNHC** | | | | | -0.14 (0.36) | | | |
| **RITAxNHC** | | | | | | -0.26 (0.17) | | |
| **Dummy AX** | | | | | | | -0.08 (0.15) | |
| **α** | | | | | | | | $0.71^{**}$ (0.25) |
| **β** | | | | | | | | 0.34 (0.20) |
| **Log Likelihood** | 155.90 | 184.69 | 185.42 | 188.99 | 184.76 | 184.88 | 184.72 | 169.18 |
| $\chi^2$ | $98.11^{***}$ | $91.14^{***}$ | $92.62^{***}$ | $92.62^{***}$ | $91.29^{***}$ | $95.53^{***}$ | $91.21^{***}$ | $93.32^{***}$ |

**Table 7.** Forecast accuracy of HFM traders and track forecasts. A trade price or forecast is correct if the probability of a hurricane making landfall in $j$ is greater than (less than) or equal to 0.5, and the hurricane makes (does not make) landfall in $j$.

| Days from | HFM | Forecast Observations | | | Fraction Correct | | | |
|---|---|---|---|---|---|---|---|---|
| Landfall | Trades | GFDL | NHC | CLP5 | HFM | GFDL | NHC | CLP5 |
| All trades | 433 | 418 | 431 | 431 | 0.84 | 0.79 | 0.81 | 0.62 |
| >5 days | 108 | 95 | 108 | 108 | 0.69 | 0.60 | 0.54 | 0.58 |
| ≤5 days | 325 | 323 | 323 | 323 | 0.89 | 0.85 | 0.90 | 0.64 |
| ≤4 days | 303 | 301 | 301 | 301 | 0.89 | 0.86 | 0.91 | 0.63 |
| ≤3 days | 270 | 268 | 268 | 268 | 0.90 | 0.89 | 0.94 | 0.63 |
| ≤2 days | 181 | 179 | 179 | 179 | 0.90 | 1.00 | 1.00 | 0.61 |
| ≤1 day | 66 | 65 | 65 | 65 | 0.98 | 1.00 | 1.00 | 0.97 |

**Table 8.** Forecast accuracy of HFM traders and track forecasts, identical forecast times. For each track forecast release, we compute the average security prices during the next six hours. Six trades for which no forecasts are available are removed from the sample.

| Days from | Observations | Fraction Correct | | | |
|---|---|---|---|---|---|
| Landfall | | HFM | GFDL | NHC | CLP5 |
| All trades | 111 | 0.94 | 0.75 | 0.77 | 0.68 |
| >5 days | 39 | 0.92 | 0.56 | 0.56 | 0.62 |
| ≤5 days | 72 | 0.94 | 0.85 | 0.89 | 0.71 |
| ≤4 days | 65 | 0.94 | 0.88 | 0.92 | 0.72 |
| ≤3 days | 52 | 0.96 | 0.94 | 0.96 | 0.71 |
| ≤2 days | 38 | 0.95 | 1.00 | 1.00 | 0.68 |
| ≤1 day | 19 | 1.00 | 1.00 | 1.00 | 0.95 |

**Figure 1.** Map of HFM landfall ranges.

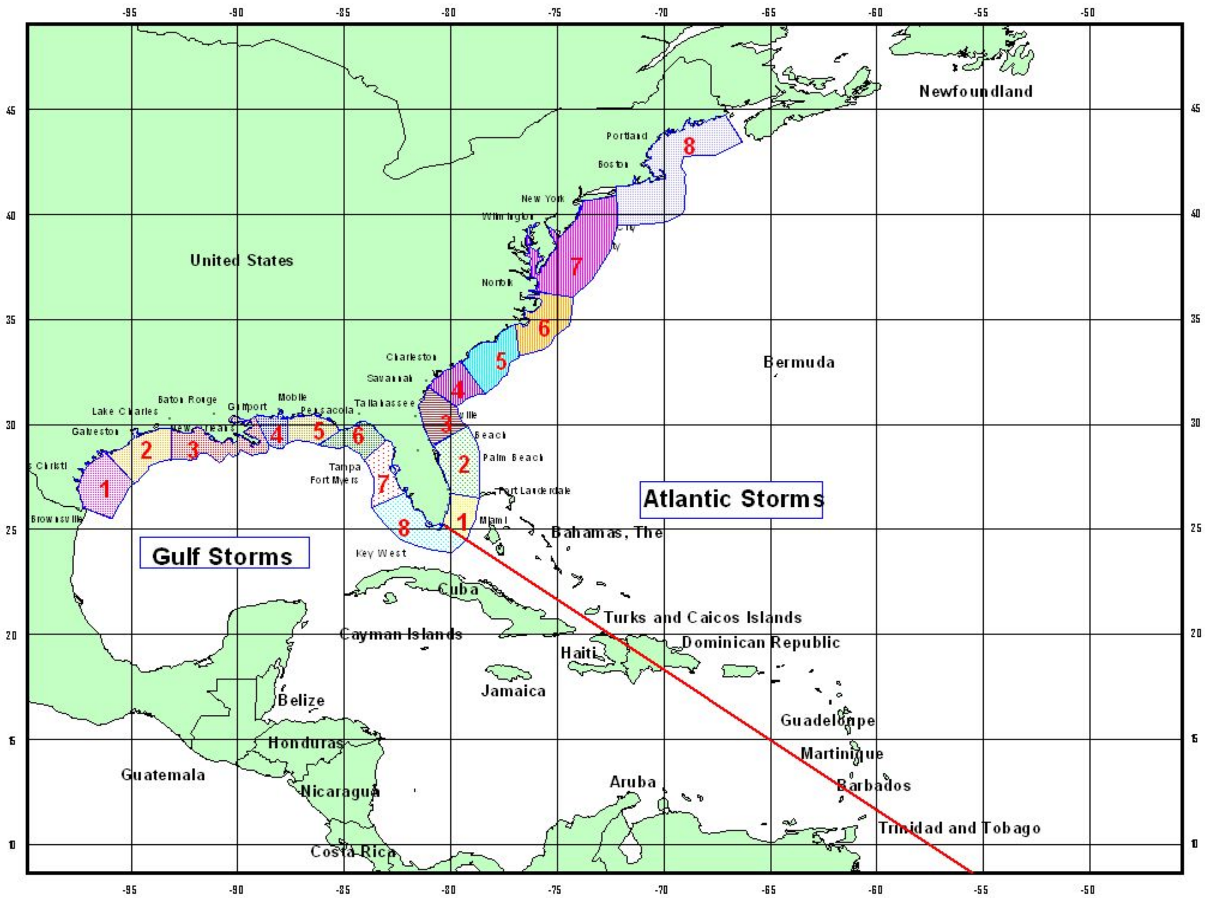


**Figure 2.** Trade-weighted average daily prices of Carolina (A5) and Expires (AX) Securities for Hurricane Ophelia, September 7-16, 2005.

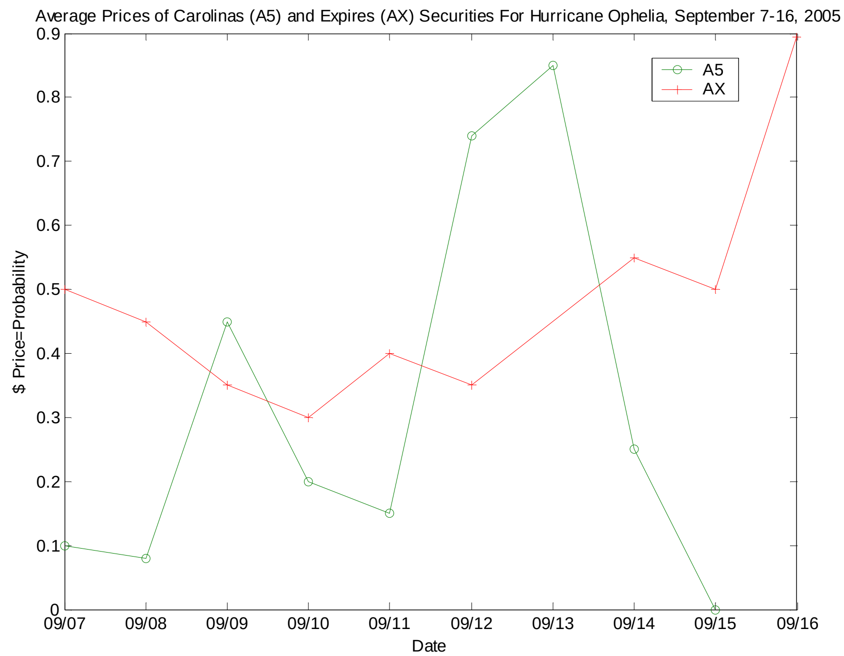



**Figure 3a.** GFDL track forecasts for Hurricane Wilma, October 17-24, 2005.

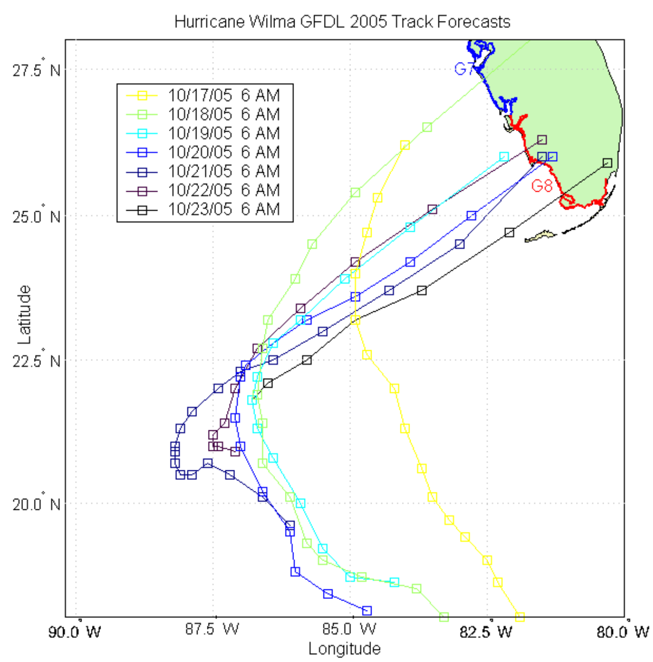


**Figure 3b.** NHC track forecasts for Hurricane Wilma, October 17-24, 2005.

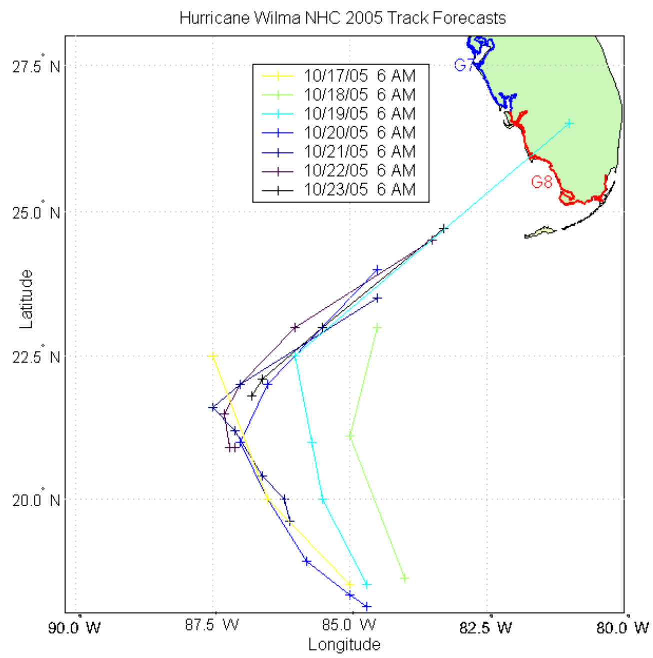



**Figure 3c.** CLP5 track forecasts for Hurricane Wilma, October 17-24, 2005.

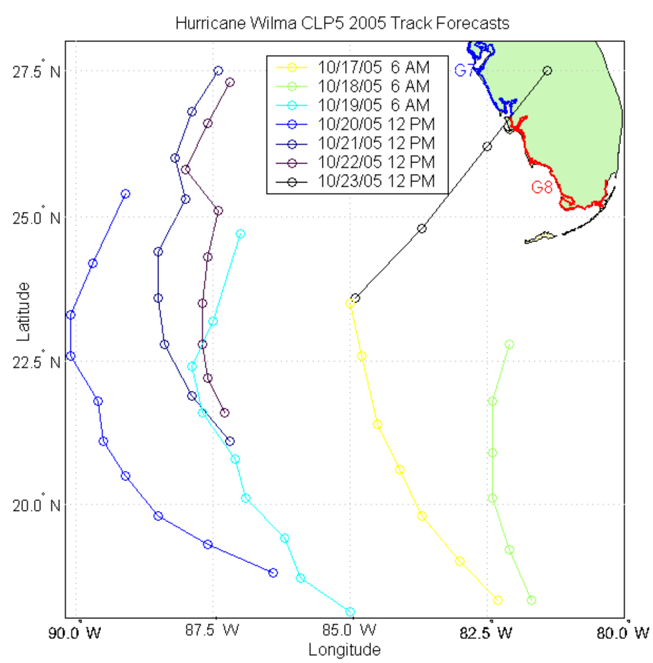


**Figure 4a.** GFDL track forecasts for Hurricane Rita, September 21-23, 2005.

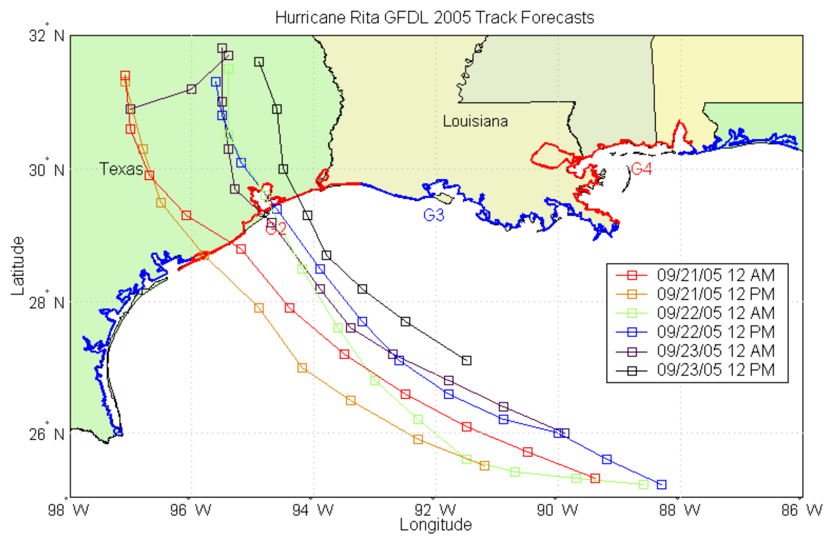



**Figure 4b.** NHC track forecasts for Hurricane Rita, September 21-23, 2005.

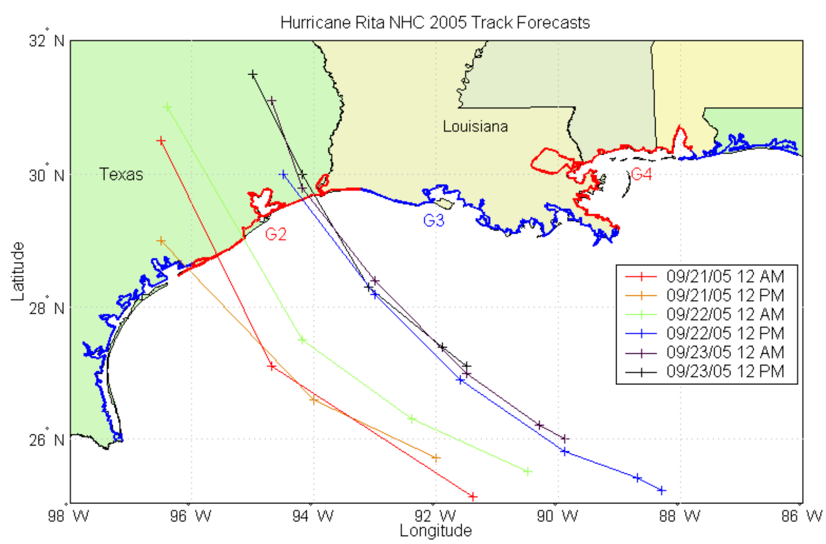


**Figure 4c.** CLP5 track forecasts for Hurricane Rita, September 21-23, 2005.

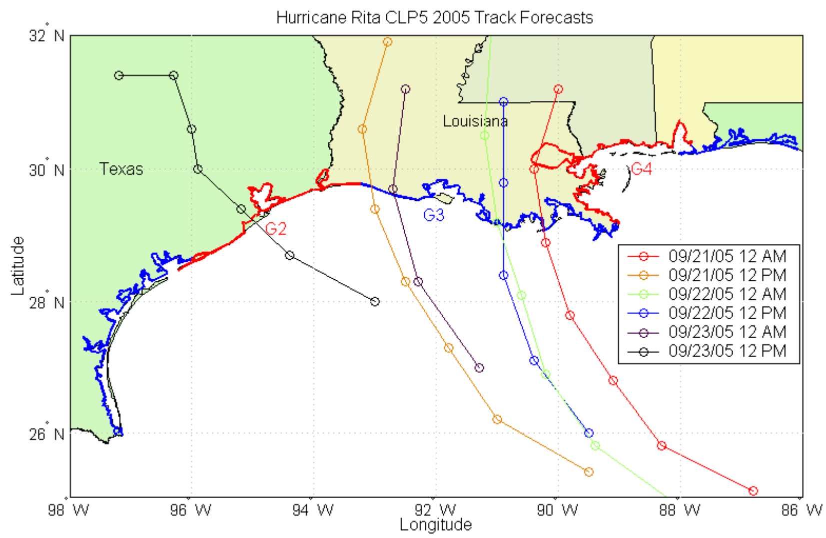



**Figure 5a.** GFDL track forecasts for Hurricane Ophelia, September 7-16, 2005.

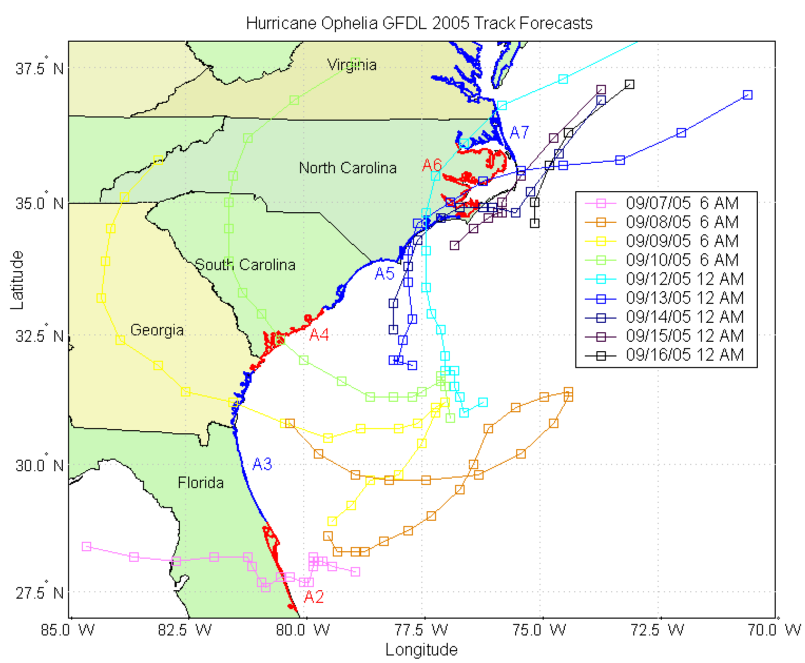


**Figure 5b.** NHC track forecasts for Hurricane Ophelia, September 7-16, 2005.

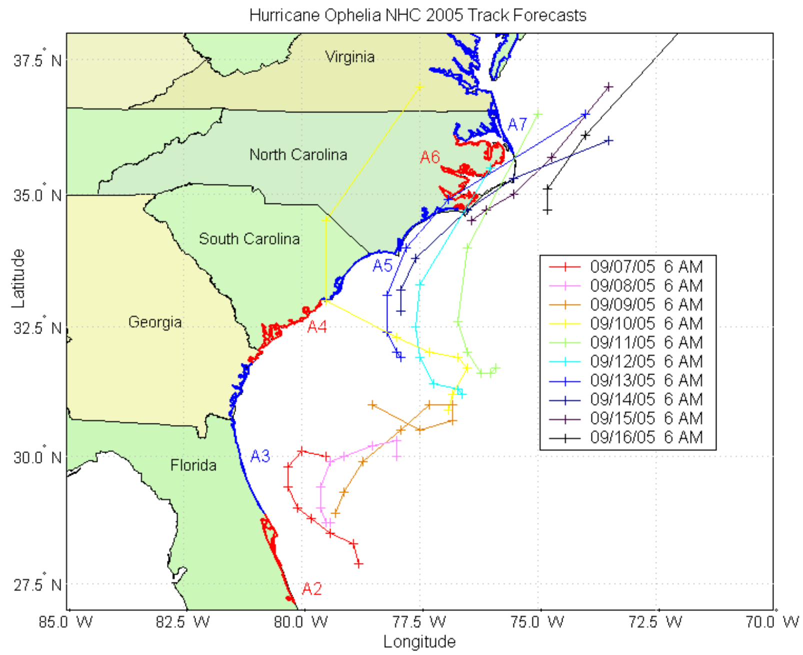



**Figure 5c.** CLP5 track forecasts for Hurricane Ophelia, September 7-16, 2005.

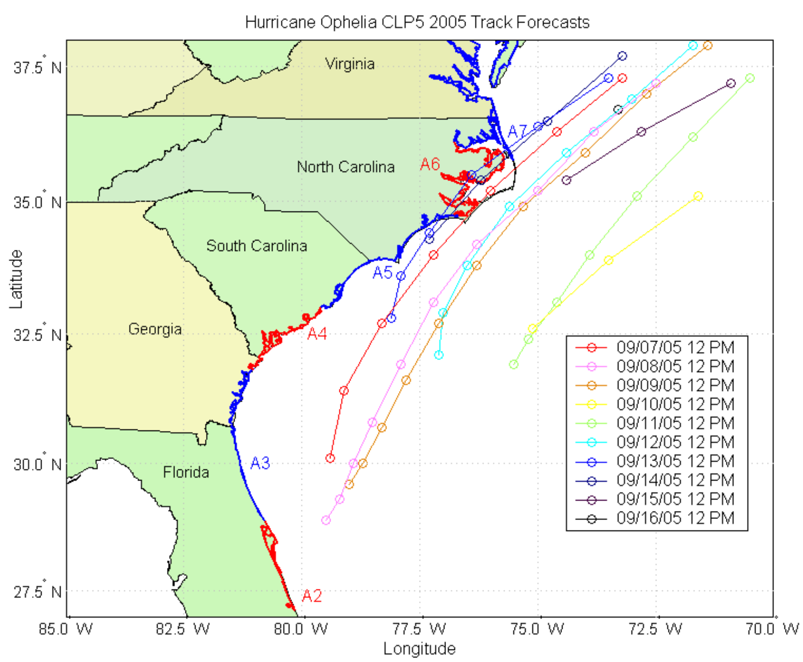


**Figure 6a.** GFDL track forecasts for Hurricane Katrina, August 27-28, 2005.

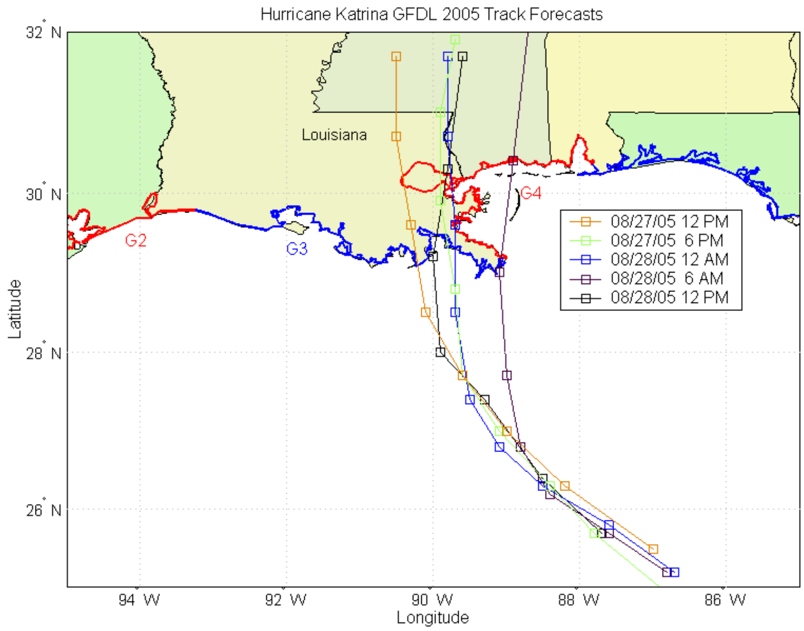



**Figure 6b.** NHC track forecasts for Hurricane Katrina, August 27-28, 2005.

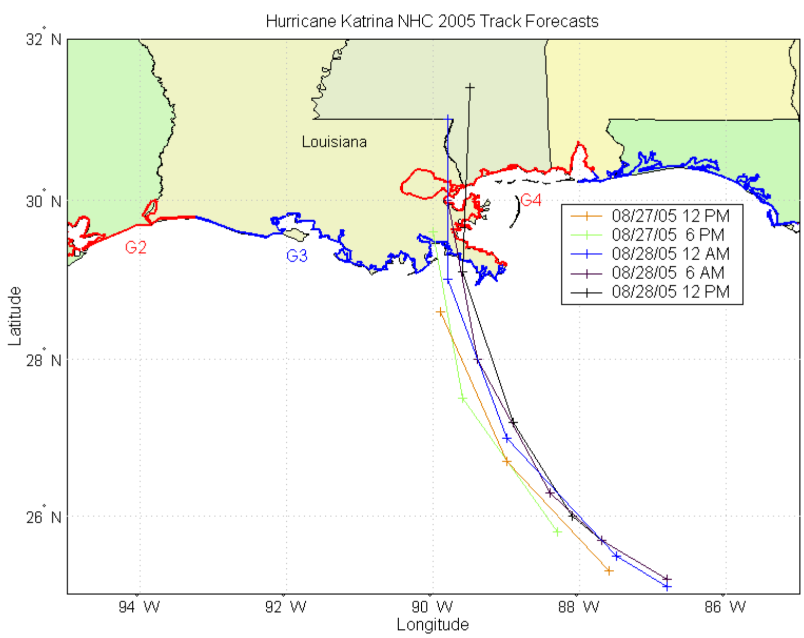


**Figure 6c.** CLP5 track forecasts for Hurricane Katrina, August 27-28, 2005.

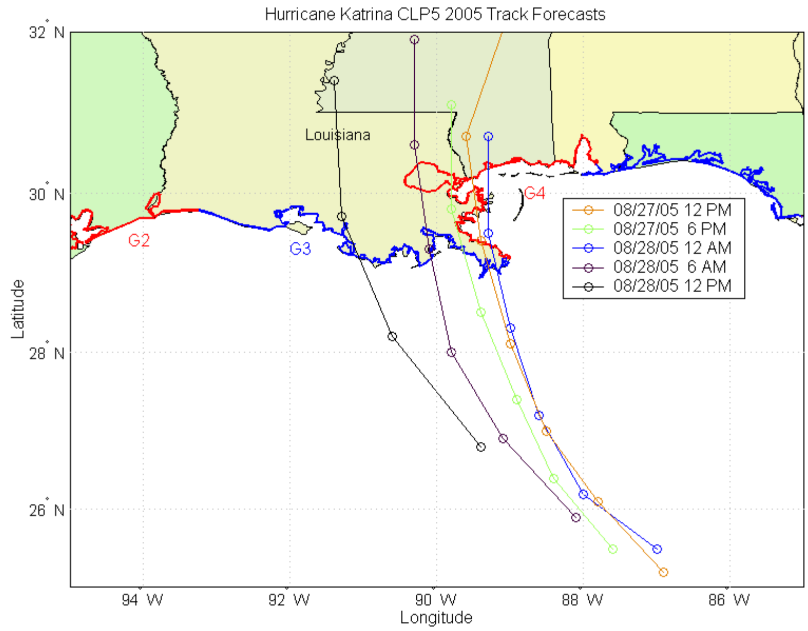



**Figure 7.** *Ex post* forecast accuracy, all storms.

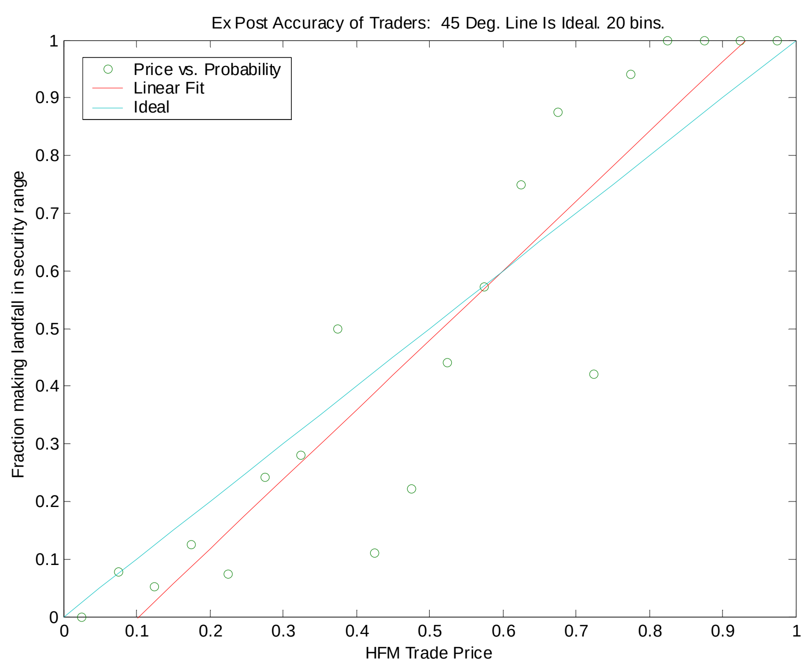


**Figure 8.** Accuracy of the normal approximation relative to the mixture of beta distributions.

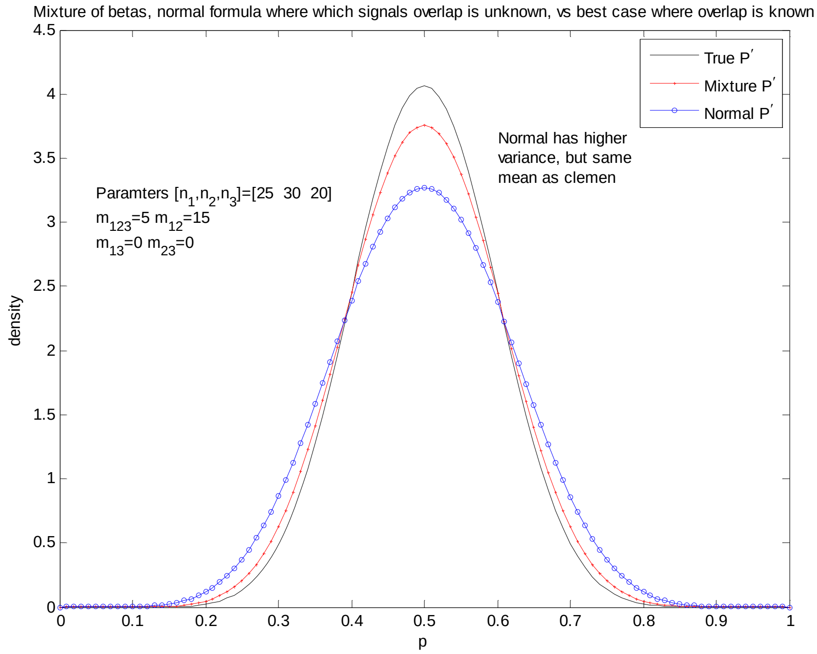
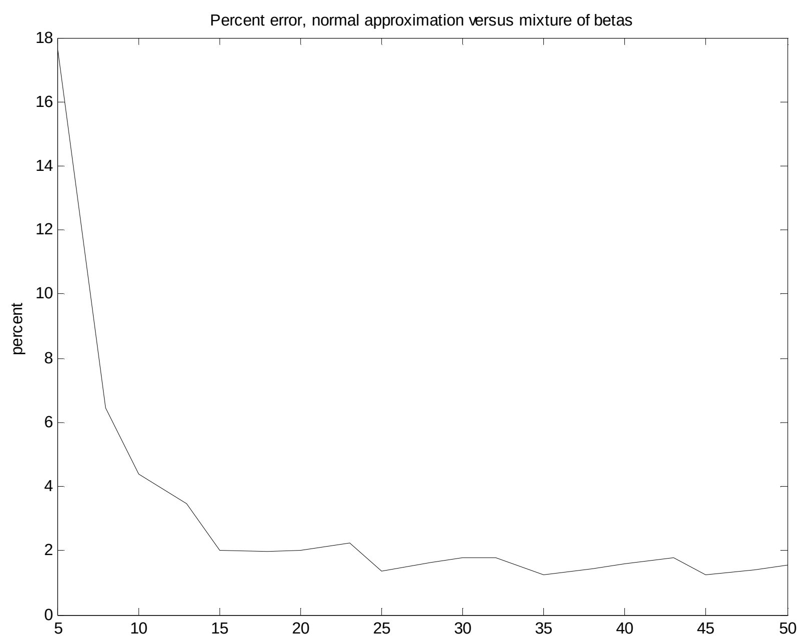



**Figure 9.** Accuracy of the normal approximation to the mixture of beta distributions. Figure varies $n=n_1+n_2+n_3+2m_{123}-m_{12}-m_{13}-m_{23}$ while keeping correlations constant. Non-monotonicity in the graph arises from changes in the ability of the mixture of betas to estimate the total number of successes as $m$ changes.



## References


Aberson, Sim D. 1998. Five-day tropical cyclone track forecasts in the North Atlantic basin, *Weather and Forecasting* 13, 1005-1015.

Baker, E. 1995. Public Response to Hurricane Probability Forecasts, *The Professional Geographer* 47, 137–147.

Bender, M.A., I. Ginis, R.E. Tuleya, Biju Thomas, and T. Marchok 2007. The Operational GFDL Coupled Hurricane–Ocean Prediction System and a Summary of Its Performance, *Monthly Weather Review* 135, 3965-3989.

Bernardo, J. and A. Smith. 2000. *Bayesian Theory*, John Wiley & Sons, Inc.

Bin, O. and S. Polasky. 2004. Effects of Flood Hazards on Property Values: Evidence Before and After Hurricane Floyd, *Land Economics* 73, 490-500.

Bolstad, W. 2004. *Introduction to Bayesian Statistics*. John Wiley & Sons, Inc.

Cameron, T. 2005. Updating Subjective Risks in the Presence of Conflicting Information: An Application to Climate Change, *The Journal of Risk and Uncertainty* 30, 63–97.

Clemen, R. 1987. Combining Overlapping Information, *Management Science* 33, 373-80.

DeGroot, M. 1970. *Optimal Statistical Decisions*. McGraw-Hill.

Dominitz, J. and C. Manski. 1997. Using Expectations Data to Study Subjective Income Expectations, *Journal of the American Statistical Association* 92, 855–867.

Ferrari, S. and F. Cribari-Neto. 2004. Beta Regression for Modelling Rates and Proportions, *Journal of Applied Statistics* 31, 799-815.

Flynn, J., P. Slovic and C. Mertz. 1994. Gender, Race, and Perception of Environmental-Health Risks, *Risk Analysis* 14, 1101-1108.

Franklin, J, C. McAdie and M. Lawrence 2003. Trends in Track Forecasting for Tropical Cyclones Threatening the United States, 1970–2001. *Bulletin of the American Meteorological Society* 84(9): 1197-1203.

Hallstrom, D. and V. Smith. 2005. Market Responses to Hurricanes, *Journal of Environmental Economics and Management* 50, 541–561.

Hanson, R. 2007. Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation, *Journal of Prediction Markets* 1, 3-15.

Jullien, B. and B. Salanie. 2000. Estimating Preferences under Risk: The Case of Racetrack Bettors, *Journal of Political Economy* 108, 503-30.

Kieschnick R. and B.D. McCullough 2003. Regression Analysis of variates on (0,1): percentages, proportions and fractions, *Statistical Modeling* 3: 193-213.

Lee, D. and E. Moretti. 2009. Bayesian Learning and the Pricing of New Information: Evidence From Prediction Markets, *American Economic Review* 99: 330-36.

Leroux, B. G. 1992. Consistent estimation of a mixing distribution, *Annals of Statistics*, 20: 1350-1360.

NHC (National Hurricane Center). 2008. Hurricane Season Tropical Cyclone Reports (all reports from 2000-6), http://www.nhc.noaa.gov/pastall.shtml.

NOAA (National Oceanic and Atmospheric Administration). 2005. *New Priorities for the 21st Century – NOAA's Strategic Plan.* Washington DC.

Oberholzer, F. and M. Mitsunari. 2006. Information Regulation: Do the Victims of Externalities Pay Attention? *Journal of Regulatory Economics* 30, 141-58.

Paulino, P. 2001. Maximum Likelihood Estimation of Models with Beta-distributed Dependent Variables, *Political Analysis* 9, 325-46.

Powell, M. and S. Aberson, 2001. Tropical Cyclone Landfall Forecasts in the Atlantic Basin (1976-2000), *Bulletin of the American Meteorological Society* 82, 2749-67.

Smith, V., D. Taylor and F. Sloan. 2001. Longevity Expectations and Death: Can People Predict Their Own Demise? *American Economic Review* 91, 1126–1134.

Smith, V. and W. Desvousges. 1988. Risk Perception, Learning and Individual Behavior, *American Journal of Agricultural Economics* 70, 1113-17.

Tetlock, P. 2004. How Efficient Are Information Markets? Evidence from an Online Exchange, University of Texas, Department of Finance, Working Paper.

Tetlock, P. 2007. Does Liquidity Affect Securities Market Efficiency? University of Texas, Department of Finance, Working Paper.

Viscusi, W. 1997. Alarmist Decisions with Divergent Risk Information, *The Economic Journal* 107, 1657-1670.

Viscusi, W. and W. Magat. 1992. Bayesian Decisions with Ambiguous Belief Aversion, *Journal of Risk and Uncertainty* 5, 371-87.

Viscusi, W. and C. O'Connor. 1984. Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers? *American Economic Review* 74, 942-956.

Wolfers, J. and E. Zitzewitz. 2004. Prediction Markets, *Journal of Economic Perspectives* 18, 107-26.

Wolfers, J. and E. Zitzewitz. 2006. Interpreting Prediction Market Prices as Probabilities, NBER Working Paper number 12200.

Zeckhauser, R. 1971. Combining Overlapping Information, *Journal of the American Statistical Association* 66, 91-2.