

**Environmental Risk and Averting Behavior: Predictive Validity of
Jointly Estimated Revealed and Stated Behavior Data**

John C. Whitehead

Appalachian State University, Boone, NC, USA

July 1, 2003 (First Revision June 17, 2004, Second Revision February 24, 2005)

Abstract. We conduct predictive validity tests using revealed and stated behavior data from a panel survey of North Carolina coastal households. The application is to hurricane evacuation behavior. Data was initially collected after Hurricane Bonnie led to hurricane evacuations in North Carolina in 1998. Respondents were asked for their behavioral intentions if a hurricane threatened the North Carolina coast during the 1999 hurricane season. Following Hurricanes Dennis and Floyd in 1999, a follow-up survey was conducted to see if respondents behaved as they intended. A jointly estimated revealed and stated behavior model indicates that the hypothetical and real evacuation behavior is based on the same choice process. Using predictions from this model with a hypothetical bias correction we find that it predicts actual evacuation behavior with small forecast error. These results suggest that stated behavior data has some degree of predictive validity.

Key words: predictive validity, stated preference, revealed preference, hurricane evacuation

JEL classification: Q54

1. Introduction

A crucial issue in the debate over the contingent valuation method and other stated preference approaches is their predictive validity. Predictive validity is the extent to which a hypothetical response accurately predicts subsequent behavior. To simplify the debate, critics of stated preference approaches argue that hypothetical questions are answered frivolously and any correlation with actual behavior is an accident. In other words, hypothetical questions have no predictive validity. Proponents argue that if hypothetical questions are incentive compatible, and all else remains constant between the survey and the time for action, hypothetical responses will closely mirror actual behavior. In other words, hypothetical questions have predictive validity.

Some degree of predictive validity of hypothetical questions is necessary for hypothetical survey data to be useful in policy analysis. If the behavioral intentions revealed during a survey have no correlation with the subsequent actual behavior, then conclusions and recommendations from the behavioral intentions data are not useful for policy. This applies to all types of stated preference data including that from contingent valuation, contingent behavior, conjoint analysis, and contingent choice studies.

Much of the empirical evidence of predictive validity has been presented within the context of the debate over the validity of the contingent valuation method for measuring passive use values (see Hanemann, 1994; Diamond and Hausman, 1994). Tests for the predictive validity of hypothetical questions have been conducted using experimental and field survey methods. A large literature exists in which actual and hypothetical willingness to pay and willingness to accept values are compared (i.e., hypothetical bias).

List and Gallet (2001) perform a meta-analysis on 29 studies published between 1972 and

2000 that focus on laboratory and field experiments of hypothetical bias. By regressing study characteristics on the ratio of hypothetical to actual values, they find that private goods generate less hypothetical bias than public goods. There are also some differences based on the incentive compatibility of the survey question. Their analysis suggests that variations in the characteristics of hypothetical questions are determinants of hypothetical bias. In general, questions based on familiar behavior (i.e., behavior that leads to use value) will generate less hypothetical bias.

In the only test of predictive validity with the contingent valuation method, Berrens and Adams (1998) are confounded by unanticipated changes in the policy after it was presented in the form of a hypothetical scenario. They suggest that the best opportunity for a test of predictive validity of stated preference methods is with stated behavior. The stated behavior predictive validity literature is relatively thin. The same-sample evidence of predictive validity from stated behavior studies suggests that survey respondents will behave in ways that they intend. Loomis (1993) re-interviews respondents eight months after a lake visitor survey interview. He finds that measures of intended length of stay during a future trip obtained from the first interview and actual length of stay during a later trip obtained from the second interview are not statistically different for similar lake conditions.

A number of studies have combined revealed and stated behavior data (e.g., Hanley, Bell, and Alvarez-Farizo, 2003). The general approach is to jointly estimate revealed behavior data and stated behavior data under different hypothetical scenarios. The benefit is the ability to obtain forecasts of behavior beyond current conditions while grounding the stated behavior data with revealed behavior data. Haener, Boxall, and Adamowicz (2001) conduct a split-sample test of predictive validity. They compare the performance of revealed behavior, stated behavior and

combined revealed and stated behavior conjoint analysis models in predicting similar site choice decisions of a holdout sample of respondents in a different recreation region. They find that the stated behavior models predict choices as accurately as revealed behavior models. But, the jointly estimated revealed and stated behavior model predicts best. They conclude by suggesting the need to test predictive validity when characteristics of the choice decision actually change relative to the base case.

Grijalva et al. (2002) conduct a same-sample predictive validity test of rock climbing trip behavior. Respondents are surveyed about their past trips and future trips under hypothetical scenarios. Following the realization of one of the hypothetical scenarios, the closure of a rock climbing area, respondents are surveyed again to determine if their stated trips are able to predict their actual trips. A conclusion from a jointly estimated revealed and stated behavior model is that the stated behavior data are predictive valid. With hypothetical closure of rock climbing areas, stated demand for climbing trips falls. When the areas are actually closed, actual changes in trip demand and welfare are similar to the hypothetical changes. However, the authors must deal with unanticipated changes in the policy after it was presented in the form of a hypothetical scenario.

Other than Loomis (1993) and Grijalva et al. (2002), no study has assessed the extent to which stated behavior data are able to predict actual behavior with the same respondents. One reason for the lack of evidence is that it is difficult to test for predictive validity. Hypothetical scenarios typically involve long-term changes in environmental quality, resource access and/or counterfactual cost changes. An additional constraint is the added expense of collecting panel data. This paper represents the first study to test the predictive validity of hypothetical averting

behavior when confronted with environmental risk. We use the same sample of respondents and assess predictive validity of hypothetical data beyond the range of historical experience. The specific behavior is a discrete choice – hurricane evacuation. We use stated behavior data obtained from a survey of North Carolina residents after Hurricane Bonnie in 1998. The stated behavior questions cover each of the five storm intensity categories in the Saffir-Simpson Hurricane Scale (Tropical Prediction Center, 1999). After Hurricanes Dennis and Floyd in 1999 we re-survey the same respondents to determine if they behaved according to predictions from the jointly estimated revealed and stated behavior model.

In the next section we describe the chronology of the hurricanes and data collection efforts. Next, evacuation behavior is compared and nonparametric tests of predictive validity are conducted. We describe a jointly estimated model of hurricane evacuation behavior using the combined revealed and stated behavior data. Then we present the results of the empirical model and comparisons of the predicted and actual evacuation behavior. A concluding section follows.

2. Data

During August of 1998 Hurricane Bonnie approached the North Carolina coast as a strong category 3 hurricane prompting evacuation orders in all eight coastal counties: Brunswick, Carteret, Currituck, Dare, Hyde, New Hanover, Onslow, and Pender. Hurricane Bonnie made landfall on the coast of North Carolina near Cape Fear, 20 miles south of Wilmington in New Hanover County (Avila, 1998). At landfall Bonnie was a low category 3 hurricane and quickly diminished to a category 1 hurricane on the Saffir-Simpson Hurricane Scale. During the next two days a weakening Bonnie made its way up the entire North Carolina coast. Much of coastal North Carolina experienced Bonnie as a tropical storm.

During January 1999 we conducted a telephone survey with a random sample of households in the North Carolina coastal counties. We completed interviews with 1029 North Carolina coastal county residents and achieved a 76% response rate. Eight-hundred ninety five respondents gave complete information and are included in the empirical study. Respondents were asked if they left their home for someplace safer during Hurricane Bonnie. We also asked respondents what they would do if a hurricane approached North Carolina during the 1999 hurricane season. About one-fifth of the survey participants were randomly assigned a hypothetical hurricane representing each category of the Saffir-Simpson Hurricane Scale. Given a hurricane watch and randomly assigned storm intensity, respondents were asked if they would evacuate their home (Appendix). If they say no, the second hypothetical question asks if they would evacuate if they were issued a voluntary evacuation order. If they say no, the third question asks if they would evacuate given a mandatory evacuation order.

During August of 1999, Hurricane Dennis approached the North Carolina coast as a category 1 hurricane, missed and stalled off the Outer Banks for two days, then landed on the northeast North Carolina coast in Dare County and quickly lost hurricane intensity (Beven, 2000). Again, portions of each of the eight North Carolina coastal counties received evacuation orders as Hurricane Dennis approached. Less than one month later, Hurricane Floyd traveled north along the Atlantic coast as a strong category 4 hurricane, weakened to a category 3 hurricane as it reached North Carolina, landed near Wilmington as a category 2 hurricane, and left the state the next day (Pasch, Kimberlain, and Stewart, 2000). Hurricane Floyd led to evacuations along the entire south Atlantic coast – which has been called the largest peacetime evacuation in United States history.

Following these storms, we revised the Bonnie survey so that it would be suitable for Hurricanes Dennis and Floyd. In January and February 2000 we attempted to re-interview the original participants from the Bonnie survey to determine who evacuated during Hurricanes Dennis and Floyd. Of the people who responded to the Bonnie survey, 66 had moved away from coastal North Carolina and 164 had either disconnected or changed phone numbers. Of the potential contacts, 565 interviews were conducted for a response rate of 68% of the Bonnie respondents. Four hundred ninety of these provided complete data.

An important difference between the 1999 and 2000 samples is the number of people who evacuated during Hurricane Bonnie. The overall evacuation rate from the 1999 survey is 26%. The Hurricane Bonnie evacuation rate for people who participated in the 2000 survey is 22%. The Hurricane Bonnie evacuation rate for the people who we were not able to contact is 31%. Another factor that helps explain who responded to the second survey is the number of years spent at the current residence. Length of time at residence increases the probability that someone responded to both surveys. This is a typical characteristic of panel surveys.¹

3. Revealed and Stated Evacuation Behavior

Evacuation behavior for the full sample is summarized in Table 1. From the first survey, 26% evacuated during Hurricane Bonnie. From the second survey, 22% evacuated during Hurricane Bonnie, 14% evacuated during Hurricane Dennis, and 34% evacuated during Hurricane Floyd. If hurricane evacuation behavior is similar across different levels of storm intensity and risk, tests of the predictive validity of stated evacuation behavior questions beyond the range of historical experience are easily satisfied and trivial. However, if storm intensity differences lead to significant differences in behavior, tests of predictive validity will be more

meaningful. The differences in revealed behavior hurricane evacuation are statistically significant at the $p = .01$ level and indicate that the three hurricanes were events with different characteristics and perceived risk levels.

Each respondent has up to three stated behavior observations: evacuation with the hypothetical storm under a hurricane watch, a voluntary evacuation order, and a mandatory evacuation order. We collapse the three stated behavior observations into a single stated behavior observation by matching the evacuation data from the hurricane watch, voluntary evacuation order, and mandatory evacuation order scenarios to the situation the respondent faced during Hurricane Bonnie. For example, if the respondent received a voluntary evacuation order during Hurricane Bonnie then we used their evacuation response to the voluntary order evacuation hypothetical scenario. About 25% of the sample stated that they would evacuate with category 1 and 2 hurricanes. Thirty-three percent, 50%, and 70% stated that they would evacuate with category 3, category 4 and category 5 hurricanes.

The comparison of stated and revealed behavior is also trivial if the averting behavior is part of a normal day to day routine when faced with a common risk. For example, a consumer who drinks several gallons of bottled water every week in response to the risk of contaminated tap water would state that they would purchase several gallons of bottled water during a future week if asked a stated behavior question. A follow-up survey would find that they, in fact, did purchase several gallons of bottled water during a typical week. In the same way, if coastal North Carolina residents are experienced with hurricanes of varying intensity and prepare for future hurricanes as part of a normal routine they will easily be able to predict their future behavior.

In the telephone surveys of coastal North Carolina residents we asked about hurricane

evacuation preparedness. Preparedness is measured by whether the respondent has developed an evacuation plan -- a recommended step of hurricane storm preparation. Only 46% of the sample had an evacuation plan following Hurricane Bonnie and slightly more, 56%, had an evacuation plan following Hurricanes Dennis and Floyd. In other words, at the time of the stated behavior survey, 54% of the sample were not familiar enough with the evacuation choice to have a plan of action in case of another storm. Further, these respondents had extensive experience with low intensity storms (i.e., categories 1 and 2) but not higher intensity storms of which the approaching Hurricane Floyd is an example. We conclude that the hurricane evacuation decision, especially with high intensity storms, is not part of the normal routine of many coastal North Carolina residents.

4. Joint Estimation of Revealed and Stated Evacuation Behavior

We jointly estimate revealed and stated evacuation behavior models using the bivariate probit model (Greene, 1999) that follows the hurricane evacuation literature (e.g., Whitehead et al. 2000)

$$\begin{aligned}
 (1) \quad & y_i^R = \alpha^R + \beta^R{}' X_i + e_i^R \\
 & y_i^S = \alpha^S + \beta^S{}' X_i + \delta' H + e_i^S \\
 & \rho = \text{corr}(e_i^R, e_i^S)
 \end{aligned}$$

where $y_i = 1$ if household i , $i = 1, \dots, 895$, chooses to evacuate and 0 otherwise, β and δ are coefficient vectors, X is a vector of risk and demographic variables, and H is a vector of storm intensity variables. Superscript R identifies the revealed behavior data and superscript S identifies the stated behavior data. The error terms, e_i , are normally distributed with zero mean and ρ is the correlation in error terms. The bivariate probit model is estimated using the LIMDEP

statistical software (Greene, 2002).

Objective hurricane risk variables include whether the respondent received voluntary or mandatory evacuation orders, whether they lived in a mobile home, and whether they lived on an island. Evacuation orders measure information received by households that suggest they are in a high hurricane risk area. Residents of mobile homes face greater risk because of the instability of the structure when faced with strong winds. Residents of islands face greater risk because of greater exposure to hurricane-force winds and storm surge. In addition to these objective risk measures we include variables that measure perceived risk from hurricane-force winds and flooding. Variables in the demographic vector include household income, education, pet ownership, race, and sex. Pet ownership acts as a constraint on evacuation behavior because shelters, most hotels, and some homes of friends and family do not allow pets. The vector of storm intensity variables includes dummy variables of hypothetical storm intensity as measured by the Saffir-Simpson Hurricane Scale. Category 3 is the omitted variable.

The independent variables for the full sample of respondents are summarized in Table 2. Fourteen percent received a mandatory evacuation order during Hurricane Bonnie. Sixteen percent of the sample received a voluntary evacuation order. Fifteen percent of the sample lives in mobile homes. Fifteen percent of the sample lives on an island. Eighty one percent perceive their wind risk to be medium or high while 42% perceive their flood risk to be so. Average annual household income is \$42,400 (1999 dollars). Average education is almost 14 years. Two-thirds of the sample has at least one pet. Eighty-four percent of the sample is white. Sixty-two percent of the sample is female.

We assess predictive validity with four revealed behavior (RB) and stated behavior (SB) models: (1) an independently estimated RB and SB model, (2) a jointly estimated RB-SB model, (3) a jointly estimated RB-SB model with common RB and SB coefficients restricted to be equal, and (4) a jointly estimated RB-SB model with common RB and SB coefficients with small differences in model (2) restricted to be equal (Table 3). The first model represents the limitations of the RB model, that it is constrained to analyzing behavior within the range of historical experience, and the limitations of the SB model, that it is based on hypothetical data. The second model recognizes that RB and SB data may be generated by the same decision process by estimating the correlation in error terms. The third model assumes that RB and SB data are generated by the same decision process by constraining the coefficients and estimating the correlation in error terms. The fourth model assumes that RB and SB data are generated by the same decision process by constraining only the coefficients that are not significantly different from each other in model 2 and estimating the correlation in error terms.

First considering the independently estimated model 1 respondents are more likely to evacuate during Hurricane Bonnie if they were given mandatory or voluntary evacuation orders, if they live in a mobile home, if they perceive medium or high flood risks, and if they are female. Respondents are less likely to evacuate if they have pets. All of the factors affecting evacuation are the same in the SB model except respondent sex does not affect the stated behavior evacuation. Among most of the statistically significant coefficient estimates, the differences in the revealed behavior and stated behavior coefficients are small -- less than 15%. The flood risk coefficient is 40% larger in the SB model. The female coefficient is 46% larger in the RB model. The constant in the RB model is 29% lower relative to the SB model. This suggests that respondents are more likely to evacuate when the decision is hypothetical.

The limitation of the RB model is that it only considers evacuation during Hurricane Bonnie, a hurricane that landed with category 2/3 intensity. The SB model contains Saffir-Simpson Hurricane Scale variables which allow consideration of evacuation beyond the range of the historical experience of Hurricane Bonnie. Relative to a category 3 hurricane, respondents are less likely to evacuate during a category 1 or 2 hurricane and more likely to evacuate during a category 4 or 5 hurricane. The coefficients suggest an increasing likelihood of evacuation as storm intensity increases. However, the constraint that the category 1 and 2 coefficients are equal cannot be rejected [$\chi^2 = .30$ (1 d.f.)]. The constraint that the category 4 and 5 coefficients are equal can be rejected [$\chi^2 = 13.52$ (1 d.f.)]. The SB model supports the univariate result from Table 1 that the likelihood of evacuation is the same for category 1 and 2 storms and increasing for categories 3, 4, and 5 hurricanes. Respondents to the stated behavior questions perceive the risk of low-intensity hurricanes to be equal and perceive increasing risk with high intensity hurricanes.

The bivariate probit model 2 results are qualitatively similar. The correlation in the error terms is 0.57 and statistically different from zero at the $p = .01$ level. The differences in the RB and SB coefficient estimates are less than 15% except for the flood risk coefficient which is 36% larger in the SB model and the female coefficient which is 71% larger in the RB model. The constant in the RB model is 33% lower relative to the SB constant.

Next we estimate bivariate probit models with parameter restrictions. The first model constrains all of the common coefficients between the RB and SB models to be equal. The qualitative results are similar to the RB model. The female coefficient remains statistically significant as in the RB model. In the constrained model, the likelihood of evacuation increases

with education. The constrained model is not statistically different from the unconstrained model [$\chi^2 = 15.62$ (12 d.f.)]. This indicates that the SB data is generated by the same choice process as the RB data. The insignificance of the coefficients on the hurricane scale category 1 and category 2 variables is a major cost to the restricted model. In contrast to the univariate and unconstrained bivariate probit results, low intensity storms have the same evacuation rates as a category 3 storm.

Finally we estimate the bivariate probit model 4 that constrains all of the common parameters between the RB and SB models to be equal except those with differences of 30% or more in coefficient estimates: the constant, flood risk and sex. The qualitative results are again similar to the RB model. The constrained model 4 is not statistically different from the unconstrained model 2 [$\chi^2 = 5.52$ (9 d.f.)] and statistically superior to the constrained model 3 [$\chi^2 = 10.10$ (3 d.f.)]. In addition coefficients on the hurricane scale category 1 and category 2 variables are statistically different from zero. In other words, low intensity storms have lower evacuation rates compared to a category 3 storm. Model 4 is the preferred model.

5. Predictive Validity

The predicted probability of evacuation is estimated from each model and compared to actual evacuations from Hurricanes Dennis and Floyd (Table 4). The evacuation probabilities are evaluated at the value of independent variables for each respondent

$$(2) \quad \hat{\pi}_{iH} = \phi(\hat{\alpha} + \hat{\beta}' X_i + \hat{\delta}_H)$$

where $H = 1, 3,$ and 4 ($\delta_3 = 0$). Forecasts for Hurricane Floyd are made assuming it is both a category 3 and 4 storm since the actual hurricane behaved differently from the generic hurricanes presented in the hypothetical scenarios. Hurricane Floyd approached North Carolina as a

category 4 storm and weakened to a category 3. The difference between the probability of evacuation and the actual behavior is the forecast error

$$(3) \quad \hat{\mu}_{iH} = \hat{\pi}_{iH} - y_{iH}$$

A t-test is used to determine if the forecast error is significantly different from zero. The sample size for these comparisons is the 490 respondents to the second survey.

Fourteen percent of respondents actually evacuated for hurricane Dennis. The SB model 1 predicts that 22% would evacuate for a category 1 storm. The forecast error is 7.7% for Hurricane Dennis (SSHS = 1). Thirty-four percent of respondents actually evacuated for Hurricane Floyd. The SB model 1 predicts that 31% would evacuate. The forecast error is -2.5% for the category 3 Hurricane Floyd and 13.8% for the category 4 Hurricane Floyd. We find statistically significant forecast error for the category 1 and 4 storms.

With the SB model 2, when the error terms from the SB and RB models are correlated, the results are similar to model 1. The bivariate probit model 3, with all twelve common parameters constrained to be equal, actually worsens the forecast accuracy of the evacuation model. Predictions for the category 1 and category 4 storms are similar to models 1 and 4. But the evacuation prediction for the category 3 Hurricane Floyd falls to 25%. The forecast error rises to -8.1% and is statistically different from zero. With the constrained RB-SB model 4, with nine of the twelve common parameters constrained to be equal, the forecast errors are similar to those from models 1 and 2.

The results do not support the paradigm that jointly estimated RB-SB models will lead to improved models of behavior when evaluated for predictive validity. However, many studies in the RB-SB literature take the model one step further by pooling the data and treating it as a

panel. With panel data, a shift in behavior from the RB data to the SB data can be tested by including a dummy variable for the SB data. When the coefficient on the dummy variable is statistically significant, it is usually interpreted as evidence of hypothetical bias. The typical correction for hypothetical bias is to set the SB variable equal to zero when making predictions about future behavior (e.g., Whitehead, 2003).

The dummy variable approach is not possible with the bivariate probit model but a similar result is found here with statistically different constant terms in model 4. Similarly to setting the SB variable equal to zero, we make forecasts with the RB constant instead of the SB constant. Since the SB constant in model 4 is larger than the RB constant, the SB data is predicting a larger number of evacuations. By using the RB constant, the evacuation forecasts will be lower than those for model 4. These results are reported in the last row of Table 4. The forecast error for Hurricane Dennis is not statistically different from zero and is about five times lower than the forecast errors from the models 1-4. The forecast errors are about four times lower when Hurricane Floyd is considered to be a category 4 storm. The forecast error is statistically significant at the $p = .10$ level. The category 3 Hurricane Floyd forecast for model 4 with the RB constant is similar to that of model 3.

The pattern of evacuation forecast errors is not unexpected for Hurricane Floyd which approached North Carolina as a strong category 4 hurricane and weakened to a category 3 hurricane before it landed with category 2 intensity. When predictions are made about Hurricane Floyd assuming that it was perceived as a category 3 storm the forecast evacuation rates are lower than the actual evacuation rate. When predictions are made about Hurricane Floyd assuming that it was perceived as a category 4 hurricane the forecast evacuation rates are above

the actual evacuation rate. An ad-hoc solution to this problem is to forecast the evacuation rate for a hurricane with intensity between category 3 and 4 storms. In other words, the coefficient on the category 4 dummy variable in model 4 is halved ($\hat{\delta}_{3.5} = 2.17 = 4.34 / 2$). The forecast probability for the category 3.5 storm is 30.8%. The forecast error of -3% is not statistically different from zero ($t = -1.50$).

6. Conclusions

Some degree of predictive validity of hypothetical questions is a necessary condition for hypothetical survey data to be useful in policy analysis. If stated behavior is positively correlated with actual behavior and of the same magnitude, then conclusions and recommendations from the stated behavior data are useful for policy. If stated behavior has little or no predictive validity, stated behavior data may be of little use for policy. Given the expense of panel data and other logistical difficulties, there is a paucity of empirical evidence about the predictive validity of hypothetical survey questions. In this paper we find some evidence that stated behavior data for hurricane evacuations is predictive valid.

We test predictive validity with jointly estimated revealed and stated behavior models. The joint estimation paradigm addresses hypothetical bias. The typical notion is that stated behavior data suffers from hypothetical bias. If stated behavior data is jointly estimated with revealed behavior data, grounding the hypothetical behavior with actual behavior, hypothetical bias may be resolved. The empirical results suggest that a jointly estimated revealed and stated behavior model with constraints on nine of twelve common parameters is most appropriate. This model predicts no better than the independently estimated SB models. This is in contrast to the results of Haener, Boxall, and Adamowicz (2001) who find that jointly estimated RB-SB models

improve on the predictions made from the independently estimated SB model. With a hypothetical bias correction the jointly estimated revealed and stated behavior model predicts evacuation for Hurricane Dennis with no statistically significant forecast error.

It is difficult to assess how well the models are able to predict evacuation for hurricane Floyd which approached North Carolina as a strong category 4 hurricane that weakened to a category 3 hurricane. Using the preferred model 4, when predictions are made about Hurricane Floyd assuming that it was perceived as a category 3 storm the forecast evacuation rate is statistically equal to the actual evacuation rate. When predictions are made about Hurricane Floyd assuming that it was perceived as a category 4 hurricane the forecast evacuation rate is well above the actual evacuation rate. With the hypothetical bias correction these results are reversed, the model does a better job of predicting evacuation with Hurricane Floyd perceived to be a category 4 hurricane. An ad-hoc solution results in a statistically accurate forecast for Hurricane Floyd.

Overall, these results suggest that stated behavior data has some degree of predictive validity. The forecast evacuation rates are reasonably close to the actual evacuation rates. In general, forecasts are increasing in storm intensity and of the same order of magnitude as actual evacuations. In the hurricane evacuation context, stated behavior data may be useful to emergency managers when issuing evacuation orders for situations beyond the range of historical significance.

One important caveat must be noted. First, achieving results that suggest that models have predictive validity is an easier task than using stated behavior data to forecast accurately. This is because testing for predictive validity provides a target for the prediction. In this research, logical decisions were made about corrections for hypothetical bias and the perceptions about the

intensity of Hurricane Floyd which led to the predictive validity conclusion. In sum, this research suggests that jointly estimated revealed and stated behavior models can achieve predictive validity. This does not suggest that jointly estimated revealed and stated behavior models will always make accurate forecasts of future behavior

Two limitations to this research must be noted. One conclusion from the List and Gallet (2001) meta-analysis is that the degree of hypothetical bias will be smaller for private goods relative to public goods. Consumption of private goods, such as hurricane evacuation, tends to be more familiar than consumption of public goods. We find that stated behavior data can be used to make reasonable predictions for the relatively familiar choice of hurricane evacuation. Similarly, Loomis (1993) and Grijalva et al. (2002) analyze familiar recreation choices. Future research should attempt to investigate the predictive validity of less familiar choices such as recreation participation with improved environmental quality amongst a sample of respondents who are non-participants with low quality.

Second, this research is limited in that the stated and revealed behaviors compared are discrete choices, decreasing the range over which behavior can vary and, perhaps, overstating the case for predictive validity. Future research on the predictive validity of stated behavior data should employ continuous measures of behavior such as recreation demand and health risk averting behavior. Also, more complex discrete choices, such as recreation site choice or transportation mode choice, might provide a more compelling context for a predictive validity study.

Acknowledgements

A previous version of this paper was presented at the AERE/ASSA Meetings, New Orleans, LA, January 2001 and the University of Central Florida, November 2001. The author thanks two anonymous referees, John Loomis and seminar participants for a number of helpful comments. This study was funded by the North Carolina Division of Emergency Management through a Memorandum of Agreement (between the North Carolina Department of Crime Control and Public Safety and East Carolina University, Regional Development Services) and the Hazard Mitigation Grant Program (HMGP 1240-0012) and by the North Carolina Sea Grant Program (NCSU 1998-0617-08).

Notes

1. One referee has reservations regarding the limitations of the data: “Generally, I do not think the context lends itself to addressing predictability of stated behavior. The dichotomous choice nature of the stated behavior is simplistic and the stated behavior following an actual event is problematic. The fact that respondents had prior experience with the choice during the ex ante survey period only magnifies the fact that respondents already had substantial familiarity and experience with this choice – a choice that is weighty and worth remembering. As the authors state, a comparison between stated and actual behavior is trivial if the choices are known. This is the case in this instance. The analysis is weak and any inference is dubious.” We attempt to address these concerns throughout the remainder of the paper.

Appendix. Stated Behavior Questions, Saffir-Simpson Hurricane Scale Category 1 Version

Please consider the following information ... hurricanes are rated on a scale of 1 to 5. Category 1 is a minimal hurricane, 2 is moderate, 3 is extensive, 4 is extreme, and 5 is a catastrophic hurricane. Bonnie was a category 3 (if asked: Fran was a 3, Bertha was a 2, and Hugo was a 4). Suppose a category 1 hurricane is approaching North Carolina. The hurricane has winds between 74 and 95 miles per hour and a storm surge about 4 to 5 feet above normal (If asked: Storm surge is the rise in sea level during a hurricane). If a Hurricane Watch is announced, would you evacuate your home to go someplace safer?

Yes (skip to next section)

No (go to next question)

If you were given a voluntary evacuation order, would you evacuate your home to go someplace safer?

Yes (skip to next section)

No (go to next question)

If you were given a mandatory evacuation order, would you evacuate your home to go someplace safer?

Yes (skip to next section)

No (go to next question)

References

- Avila, Lixion A. (1998), Preliminary Report: Hurricane Bonnie, 19-30 August 1998. National Hurricane Center, <http://www.nhc.noaa.gov/1998bonnie.html>, October.
- Berrens, R., and R. Adams (1998), "Using Contingent Valuation to Design Fee Hunting Programs: Oregon Pheasant Hunting Revisited." *Human Dimensions of Wildlife* **3** (3), 24-38.
- Beven, Jack (2000), Preliminary Report: Hurricane Dennis 24 August – 7 September 1999. National Hurricane Center, http://www.nhc.noaa.gov/1999dennis_text.html, January.
- Diamond, Peter A. and Jerry A. Hausman (1996) "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* **8** (4) 45-64.
- Freund, John E., and Ronald E. Walpole (1980), *Mathematical Statistics*, Third Edition, Englewood Cliffs, NJ: Prentice-Hall.
- Greene, William H. (2002), *LIMDEP Version 8.0 User's Manual*, Plainview, NY: Econometric Software, Inc.
- Greene, William H. (1999), *Econometric Analysis*, Third Edition. Upper Saddle River, NJ: Prentice Hall.
- Grijalva, Therese, Robert P. Berrens, Alok K. Bohara, and W. Douglass Shaw (2002), "Testing the Validity of Contingent Behavior Trip Responses," *American Journal of Agricultural Economics* **84** (2), 401-414.
- Hanemann, W. Michael (1994) "Valuing the Environment through Contingent Valuation," *Journal of Economic Perspectives* **8** (4), 19-44.
- Haener, Michel K., Peter C. Boxall, and Wictor L. Adamowicz (2001), "Modeling Recreation Site Choice: Do Hypothetical Choices Reflect Actual Behavior?" *American Journal of*

- Agricultural Economics* **83** (3), 629-642.
- Hanley, Nick, David Bell, and Begona Alvarez-Farizo (2003), "Valuing the Benefits of Coastal Water Quality Improvements Using Contingent and Real Behaviour," *Environmental and Resource Economics* **24** (3), 273-285.
- List, John A., and Craig A. Gallet (2001), "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Values?" *Environmental and Resource Economics* **20** (3), 241-254.
- Loomis, John B. (1993), "An Investigation in the Reliability of Intended Visitation Behavior," *Environmental and Resource Economics* **3** (2), 183-191.
- Pasch, Richard J., Todd B. Kimberlain, and Stacy R. Stewart (2000), Preliminary Report: Hurricane Floyd, 7– 17 September 1999. National Hurricane Center, http://www.nhc.noaa.gov/1999floyd_text.html, January.
- Tropical Prediction Center (1999), The Saffir-Simpson Hurricane Scale. National Hurricane Center, <http://www.nhc.noaa.gov/aboutsshs.html>, July 14.
- Whitehead, John C., Bob Edwards, Marieke Van Willigen, John R. Maiolo, Kenneth Wilson, and Kevin T. Smith (2000), "Heading for Higher Ground: Factors Affecting Real and Hypothetical Hurricane Evacuation Behavior," *Environmental Hazards* **2** (3), 133-142.
- Whitehead, John C. (2003), "One Million Dollars per Mile? The Opportunity Costs of Hurricane Evacuation," *Ocean and Coastal Management*, **46** (11-12), 1069-1083.