University of Wisconsin-Madison

IRP Discussion Papers

Charles F. Manski

THE USE OF INTENTIONS DATA TO PREDICT BEHAVIOR: A Best-Case Analysis

DP #880-89

Institute for Research on Poverty
Discussion Paper no. 880-89

THE USE OF INTENTIONS DATA TO PREDICT BEHAVIOR: A BEST-CASE ANALYSIS

Charles F. Manski
Department of Economics
and
Institute for Research on Poverty
University of Wisconsin-Madison

May 1989

This work is supported by National Science Foundation Grant SES-8808276 and by a grant from the U.S. Department of Health and Human Services to the Institute for Research on Poverty. Some of the ideas developed here originated in joint research with Daniel Nagin, undertaken around 1980. A note titled "Behavioral Intentions and Revealed Preferences" was circulated informally at that time. More recently, I have benefited from useful discussions with Rob Mare and Jim Walker.

Abstract

In surveys individuals are routinely asked to predict their future behavior, that is, to state their intentions. The use of intentions data to predict behavior has been controversial. This paper determines the information content of intentions under the "best-case" assumption that individuals respond as would persons with rational expectations. It is found that intentions data bound but do not identify the probability that a person will behave in a given way. Some mixed empirical evidence on consistency with the bounds is presented. Two alternatives to traditional intentions questions are considered. These are "probability forecasts" and "forced-choice" questions.

THE USE OF INTENTIONS DATA TO PREDICT BEHAVIOR: A BEST-CASE ANALYSIS

"INTENTION implies little more than what one has in mind to do or bring about," Webster's Ninth New Collegiate Dictionary (1985, p. 629)

1. INTRODUCTION

In surveys individuals are routinely asked to predict their future behavior, that is, to state their intentions. The fertility question asked female respondents in the June 1987 Supplement to the Current Population Survey (CPS) is an example:

> Looking ahead, do you expect to have any (more) children?
> Yes No Uncertain
>
> (U.S. Bureau of the Census, 1988)

Another example is the set of schooling-work questions asked in fall 1973 of respondents to the National Longitudinal Study of the High School Class of 1972 (NLS72).

What do you expect to be doing in October 1974?

(Circle one number on each line)

	Expect to be doing	Do not expect to be doing
Working for pay at a full-time or part-time job	1	2
Taking vocational or technical courses at any kind of school or college	1	2
Taking academic courses at a two-year or four-year college	1	2
On active duty in the Armed Forces (or service academy)	1	2
Homemaker	1	2

(Riccobono et al., 1981)

Social scientists seeking to predict behavior differ in the use they make of such questions. Economists, observing the high frequency of discrepancy between stated intentions and subsequent behavior, generally ignore the responses. Their practice has been to use data on past behavior, filtered through econometric models of decision making, to predict future behavior. See, for example, the predictions of schooling-work behavior in Manski and Wise (1983, Chap. 7).

Demographers often use intentions data to predict behavior. A common practice is to make intentions the intermediate variable in a path model. One equation explains intentions as a function of respondent background. A second explains behavior as a function of respondent background and intentions. See, for example, the analysis of fertility intentions in Westoff and Ryder (1977).

Neither of these practices is well grounded. It is reasonable for economists to ignore stated intentions only if these data contain no information about future behavior. The fact that intentions and behavior often diverge does not, however, imply that the former are uninformative about the latter. At the same time, the demographic literature offers no compelling reason to think that path models appropriately express the relationship between intentions and behavior.

To use intentions data fruitfully, we need to understand how individuals respond to the survey questions asked of them. This paper explores the implications of the economic hypothesis that individuals respond as would persons with rational expectations.

The rational-expectations hypothesis is interesting not because it is necessarily realistic but because it places an upper bound on the behavioral information contained in intentions data. Under rational expectations, stated intentions are best point predictors of subsequent behavior. Thus this paper studies a "best-case" scenario for the use of intentions data to predict behavior.

The paper examines intentions questions that concern binary choices, as do the fertility and schooling-work questions cited earlier. Intentions data about binary choices are easier to interpret than are responses to quantitative questions. (For example, the CPS fertility supplement asks "How many more do you expect to have?") So again the focus is on a best-case scenario.

Even within the restricted world of rational expectations and behavior based on binary choices, the interpretation of intentions data requires care. The rational-expectations hypothesis does not imply that intentions and behavior always coincide. The two may diverge whenever the information available to the respondent at the time of the survey is more limited than the information he or she will possess at the later time when behavior is determined.

The discrepancies between intentions and behavior can be large. Suppose, for example, that all of the NLS72 respondents say that they expect to be working in October 1974. Suppose that only 50 percent subsequently do so. We shall show that this pattern of responses is consistent with rational expectations.

It does not imply that respondents wrongly predict their future behavior.

On the other hand, the discrepancies between intentions and behavior cannot be arbitrarily large. Consider again the NLS72 questions. If respondents have rational expectations and if an auxiliary condition holds, the fraction who do work must be within 1/2 of the fraction who say they intend to do so. Thus intentions data do contain some information about behavior.

The foregoing findings are developed in Section 2, which shows that intentions data bound but do not identify the probability that a person will behave in a given way. Section 3 applies these bounds to the problem of testing the rational-expectations hypothesis and presents some mixed empirical evidence. Section 4 considers "probability forecasts" as an alternative to the traditional survey question on intentions. Section 5 discusses "forced-choice" questions, which are distinct from but sometimes confused with intentions questions. Section 6 offers conclusions.

Although the substantive concern of this paper is the use of intentions data, most of the analysis applies to a larger class of prediction questions asked in surveys. Individuals are often asked to make a point prediction of some future binary event, not necessarily their own future behavior. For example, economists are often asked to predict whether the unemployment rate will rise or fall. All of the analysis in Sections 2 through 4 applies to such questions. Only the discussion of Section 5 is specific to questions regarding the respondent's own behavior.

2. BOUNDS ON BEHAVIOR IMPLIED BY BINARY INTENTIONS DATA

2.1. The Survey Question and the Rational-Expectations Response

In Sections 2 and 3 we suppose that a person is asked to make a point prediction of some binary choice. That is, a yes/no answer is requested, as in each of the NLS72 schooling-work questions. Section 4 will consider situations in which the respondent is allowed to express uncertainty about his or her future, as in the CPS fertility question.

Let i and y be zero-one indicator variables denoting the survey response and future behavior respectively. Thus $i = 1$ if the person responds "yes" to the survey question and $y = 1$ if his behavior turns out to satisfy the property of interest.

To form his response, a person with rational expectations would begin by recognizing that his behavior will depend in part on conditions known to him at the time of the survey and in part on events that have not yet occurred. Let s denote the information available to the respondent at the time of the survey. Let z denote the events that have not yet occurred but which will affect his future behavior. Thus z represents uncertainty which will be resolved between the time of the survey and the time at which the behavior is determined. The behavior y is a function of the pair (s,z) and so may be written $y(s,z)$.

Let $P_z|s$ denote the objective probability distribution of z conditional on s. Let $P(y|s)$ denote the objective distribution of y conditional on s. The event $y = 1$ occurs if and only if the realization of z is such that $y(s,z) = 1$. Hence

$$(1) \quad P(y=1|s) \quad = \quad P_z[y(s,z)=1|s].$$

The content of the rational-expectations hypothesis is that, at the time of the survey, the respondent knows $y(s,*)$ and $P_z|s$; hence he knows $P(y=1|s)$. (It does not suffice for the respondent to have a subjective distribution for z, from which he derives a subjective distribution for y. The rational-expectations hypothesis assumes knowledge of the actual stochastic process generating z.)

The respondent to the survey question provides a point prediction of his behavior. The wording of the question varies in practice. The respondent may be asked to state what he "expects," "intends," or "is likely" to do. Variations on wording may legitimately affect a respondent's interpretation of the question. Continuing our best-case analysis, however, we shall suppose that they do not. In particular, we shall maintain the hypothesis that the respondent provides the best point prediction of his behavior under a symmetric loss function. Then the response i is the more likely of the two events $y = 0$ and $y = 1$. Thus the condition

$$(2)\quad i = 1 \Rightarrow P(y=1|s) \geq 1/2$$
$$i = 0 \Rightarrow P(y=1|s) \leq 1/2$$

expresses the information on behavior contained in the stated intentions.

2.2. Prediction of Individual Behavior Conditional on Intentions

Now consider a researcher who wishes to use intentions data to predict the behavior of a given respondent. The researcher observes the survey response i. Typically he observes only a subset of the information s available to the respondent. Let x denote the observed component of s.

Suppose that the researcher wishes to predict the behavior y conditional on the observed variables x and i. Then he would like to learn the probability $P(y=1|x,i)$. Intentions data do not identify $P(y=1|x,i)$. They do, however, imply a bound.

Let $P_s|xi$ denote the probability distribution of s conditional on the observed pair (x,i). In general,

$$(3)\quad P(y=1|x,i) = \int P(y=1|s)dP_s|xi.$$

It follows directly from this and from (2) that

$$(4)\quad P(y=1|x,i=0) \leq 1/2 \leq P(y=1|x,i=1).$$

This bound of width 1/2 expresses all the information about $P(y|x,i)$ contained in intentions data. Note that the position of the bound does not vary with x.

The foregoing implies that some familiar models for $P(y|x,i)$ are not consistent with the rational-expectations hypothesis. Consider first the binary logit model

$$(5) \quad P(y=1|x,i) = \frac{\exp(x\beta+\gamma i)}{1 + \exp(x\beta+\gamma i)},$$

where (β,γ) are parameters. This model has the property

$$(6) \quad \begin{array}{lll} x\beta + \gamma i < 0 & \Rightarrow & P(y=1|x,i) < 1/2 \\ x\beta + \gamma i = 0 & \Rightarrow & P(y=1|x,i) = 1/2 \\ x\beta + \gamma i > 0 & \Rightarrow & P(y=1|x,i) > 1/2. \end{array}$$

Condition (6) is consistent with (4) only if (x,β,γ) satisfies the special property $x\beta \leq 0 \leq x\beta + \gamma$.

This problem is, of course, not limited to the logit model. It is a characteristic of any model which attempts to explain y as a function of a linear index $x\beta+\gamma i$.

2.3. Prediction Not Conditional on Intentions

Often a researcher wants to predict the behavior of a nonsampled member of the population from which the survey respondents were drawn. Intentions data are available only for the sampled individuals. But some background variables x may be observed for the entire population. In this setting, one may want to predict behavior conditional on these x. Then the quantity of interest is $P(y=1|x)$.

The bound (4) implies a bound on $P(y=1|x)$. Observe that

$$(7) \quad P(y=1|x) \equiv P(y=1|x,i=0)P(i=0|x) + P(y=1|x,i=1)P(i=1|x).$$

It follows from (4) and (7) that

$$(8) \quad P(i=1|x)/2 \leq P(y=1|x) \leq 1/2 + P(i=1|x)/2.$$

This bound, unlike (4), does vary with x.

The bound (8) is useful in practice if the quantity $P(i=1|x)$ can be estimated consistently from the sample data. Under the maintained assumption of random sampling, this is generally possible. If x is discrete, a suitable estimate is the fraction of respondents with observed characteristics x who answer "yes" to the survey question. If x is continuous, nonparametric regression methods may be applied. See, for example, Prakasa Rao (1983), Bierens (1987), or Manski (1988). Thus (8) provides an estimable bound of width 1/2 on $P(y=1|x)$.

Analyses of intentions data sometimes presume that the sharp relationship

$$(9) \quad P(i=1|x) = P(y=1|x)$$

should hold. Deviations from this equality are considered "inconsistencies" in need of explanation. For example, Westoff and Ryder (1977) state:

> The question with which we began this work was whether reproductive intentions are useful for prediction. The basic finding was that 40.5 percent intended more, as of the end of 1970, and 34.0 percent had more in the subsequent five years . . . . In other words, acceptance of 1970 intentions at face value would have led to a substantial overshooting of the ultimate outcome. (p. 449)

Seeking to explain the observed "overshooting" of births, the authors state:

> one interpretation of our finding would be that the respondents failed to anticipate the extent to which the times would be unpropitious for childbearing, that they made the understandable but frequently invalid assumption that the future would resemble the present--the same kind of forecasting error that demographers have often made. (p. 449)

More recent demographic work maintains the presumption that deviations from (9) require explanation. See, for example, O'Connell and Rogers (1983).

The rational-expectations hypothesis implies that (9) should hold in one very special case; that in which future behavior depends only on the information s available at the time of the survey. In this case, the respondent can forecast his future behavior with certainty. So i always equals y.

In the nondegenerate case when future events z partially determine behavior, the rational-expectations hypothesis does not imply (9). Let 1[*] denote the indicator function taking the value one if condition [*] is satisfied and zero otherwise. It follows from (2) that

$$(10) \quad P(i=1|x) = \int 1[P(y=1|s)>1/2]dP_s|x,$$

provided only that the event $P(y=1|s) = 1/2$ occurs with probability zero conditional on x. On the other hand,

$$(11) \quad P(y=1|x) = \int P(y=1|s)dP_s|x.$$

The right-hand sides of (10) and (11) are not generally equal.

A simple example makes the point forcefully. Suppose that $P(y=1|s) = .51$ for all values of s. Then $P(y=1|x) = .51$ but $P(i=1|x) = 1$.

3. CONSISTENT BOUNDS TESTS OF THE RATIONAL-EXPECTATIONS HYPOTHESIS

3.1. The No-Aggregate-Shocks Condition

In introducing the rational-expectations hypothesis, we emphasized its "best-case" property; it makes intentions a best predictor of behavior. One may also wish to know whether the rational-expectations hypothesis describes the way people really respond to intentions questions. The bounds (4) and (8) suggest simple tests of this hypothesis.

Suppose that one observes x and i on an initial survey of a random sample of the relevant population and observes y on a later resurvey. Assume, for simplicity, that x is discrete. Then $P(i=1|x)$, $P(y=1|x)$, and $P(y=1|x,i)$ can be estimated by the corresponding sample frequencies, and one may check whether the estimates satisfy the bounds.

These bounds tests are consistent provided that the estimates for $P(i=1|x)$, $P(y=1|x)$, and $P(y=1|x,i)$ are consistent. Random sampling ensures that the estimate for $P(i=1|x)$ is consistent. The estimates for $P(y=1|x)$ and $P(y=1|x,i)$ are consistent as long as the realizations of the future events z are not too dependent across the population. This auxiliary condition is sometimes referred to as the absence of "aggregate shocks."

To see the problem that dependence can cause, consider the extreme case in which a single event z is drawn from the distribution $P_z|s$, and all the people characterized by s realize this event. Then $P(y=1|s)$ continues to be given by equation (1) but, conditioning on s, the realized frequency of the event $y = 1$ can be only zero or one, depending on what single realization of z is drawn.

3.2. Empirical Evidence from the NLS72

The NLS72 offers an opportunity to perform the tests. The intentions questions quoted at the beginning of Section 1 were followed by behavior questions asked in the fall of 1974:

What were you doing the <u>first week of October 1974?</u>
(Circle as many as apply)

Working for pay at a full-time or part-time job	1
Taking academic courses at a two- or four-year college	2
Taking vocational or technical courses at any kind of school or college	3
On active duty in the Armed Forces (or service academy)	4
Homemaker	5
Temporary lay-off from work, looking for work, or waiting to report to work	6

These questions about behavior correspond closely, although not perfectly, to the intentions questions asked a year earlier. One difference is that a "temporary lay-off" question was added to the 1974 survey. A second is that the instructions call for respondents to "circle as many as apply" rather than to "circle one number on each line." A third difference is that the respondents to the intentions questions were asked to forecast their behavior in October 1974, whereas the behavior questions concern the first week of that month. These distinctions will be ignored here, although it is possible that they are germane.

Table 1 presents findings for the case in which x gives the respondent's sex. The probabilities $P(i=1|x)$, $P(y=1|x)$, and $P(y=1|x,i)$ are estimated by corresponding sample frequencies. For example the estimate of $P(work=1|x=male,i=0)$ is based on the 2546 males who said they did not expect to work; the fraction of this group who reported working a year later was .42.

Table 1

Consistency of Schooling-Work Intentions Stated in Fall 1973 with Behavior in October 1974

BOUND (4)

Behavior		Males: Number of Observations	Males: Estimate of P(y=1|x,i)	Females: Number of Observations	Females: Estimate of P(y=1|x,i)
Work	i=1	7143	.80	7439	.71
	i=0	2546	.42	2688	.31
Voc-Tech	i=1	1929	.16	1706	.15
	i=0	6972	.03	7575	.03
Academic	i=1	4829	.67	4320	.68
	i=0	4509	.06	5320	.06
Military	i=1	966	.67	158	.42
	i=0	7951	.012	8938	.002
Homemaker	i=1	246	.09	3626	.64
	i=0	8464	.004	5753	.08

BOUND (8)

Behavior	Males: Estimates of P(i=1|x)	Males: Estimates of Bound	Males: Estimates of P(y=1|x)	Females: Estimates of P(i=1|x)	Females: Estimates of Bound	Females: Estimates of P(y=1|x)
Work	.64	[.32-.82]	.67	.65	[.33-.83]	.58
Voc-Tech	.17	[.09-.59]	.05	.15	[.08-.58]	.05
Academic	.43	[.22-.72]	.33	.38	[.19-.69]	.30
Military	.09	[.05-.55]	.08	.014	[.007-.507]	.009
Homemaker	.02	[.01-.51]	.005	.32	[.16-.66]	.27

NOTE: The number of observations is not the same across questions because some respondents did not answer some questions. For example, 9689 males (i.e., 7143+2546) answered the work intentions question while 8917 (i.e., 966+7951) answered the military question.

The table shows mixed findings. The male and female responses to the "work" and "academic" questions satisfy bounds (4) and (8). So do the male responses to the "military" question and the female responses to the "homemaker" question. The female responses to the "military" question satisfy the bounds except for a modest violation of bound (4) by those stating $i = 1$.

On the other hand, the responses of both sexes to the "voc-tech" question and the male responses to the "homemaker" question violate the bounds substantially. Respondents who say they expect to take voc-tech courses subsequently do so only 15 or 16 percent of the time. Males who say they expect to be homemakers later report themselves as such only 9 percent of the time.

It is possible, but unlikely, that these findings reflect sampling variation. The estimates for $P(i=1|x)$ and $P(y=1|x)$ are based on samples of roughly 10,000 observations. Those for $P(y=1|x,i)$ are based on samples whose minimum size is 158 and which generally have several thousand observations.

The estimates for $P(i=1|x)$ are from random samples and therefore are extremely precise. The estimates for $P(y=1|x)$ and $P(y=1|x,i)$ must also be precise unless the realizations of z are strongly dependent across the population of NLS72 respondents. It is not obvious, however, what aggregate shock occurring between fall, 1973 and October, 1974 might generate such dependence.

3.3. Empirical Evidence from the National Fertility Survey

The National Fertility Survey of 1970 offers an additional opportunity to test the rational-expectations hypothesis. Westoff and Ryder (1977) compare the responses on fertility intentions in the 1970 survey with subsequent fertility reported in a 1975 resurvey. The findings are mostly but not entirely positive.

Of those woman who say they intend to have more children, 66 percent do so. Of those who say they do not intend to have more children, 12 percent do. These results satisfy the bounds.

On the other hand, a bound is violated by the responses to a question on the timing of births. Those women who stated that they intend to have more children were asked the further question, "Do you intend to have your next child within two years from now?" Of those women who responded affirmatively (negatively), 41 (29) percent did subsequently have a child within the next two years. Thus, those answering affirmatively violate bound (4).

4. PROBABILITY FORECASTS

Survey researchers using the traditional intentions question ask the respondent to make a point prediction of his future behavior. One could instead ask for a probability forecast. Under the rational-expectations hypothesis, this amounts to asking the respondent for the value of $P(y=1|s)$.

Probability forecasts are much more informative about behavior than are point forecasts. Intentions data reveal only the bounds (4) and (8) on $P(y|x,i)$ and $P(y|x)$ respectively. Probability forecasts identify $P(y|x,i)$ and $P(y|x)$ themselves. To see this, observe that

$$(12a)\quad P(y|x,i=0) = E[P(y|s)\,|\,x,i=0] = E[P(y|s)\,|\,x,P(y|s)\le 1/2]$$

$$(12b)\quad P(y|x,i=1) = E[P(y|s)\,|\,x,i=1] = E[P(y|s)\,|\,x,P(y|s)\ge 1/2]$$

$$(12c)\quad P(y|x) = E[P(y|s)\,|\,x].$$

If x is discrete, the right-hand-side expression in (12a) can be estimated by the sample average of the probability forecasts $P(y=1|s)$ among the subsample of respondents who are characterized by x and who report $P(y=1|s) \le 1/2$. The right hand sides of (12b) and (12c) can be estimated by analogous sample averages. If x is continuous, nonparametric regression methods can be applied.

Survey researchers very rarely ask for probability forecasts. The prevailing view among them seems to be that people are able to respond sensibly to traditional intentions questions but not to probability questions. It is unclear whether this view has an empirical justification.

Survey instruments such as the CPS try to move in the direction of probability forecasts by asking the respondent to respond "yes," "no," or "uncertain" to the intentions question. The usefulness of this type of question depends critically on how respondents interpret the three response categories. In the best case, there is consensus among respondents that

(13) "no" $\Rightarrow$ $P(y=1|s) \leq \kappa_1$

"uncertain" $\Rightarrow$ $\kappa_1 \leq P(y=1|s) \leq \kappa_2$

"yes" $\Rightarrow$ $P(y=1|s) \geq \kappa_2$,

where $0 \leq \kappa_1 \leq \kappa_2 \leq 1$ and where κ_1 and κ_2 are known to the researcher. On the other hand, it may be that the values κ_1 and κ_2 vary across respondents in a manner not understood by the researcher. If so, the responses to the CPS question are difficult to interpret.

5. FORCED-CHOICE QUESTIONS

A "forced-choice" question requires the respondent to decide what future behavior he would choose if he had to commit himself now. Forced-choice and intentions questions are sometimes confused. Voter pre-election surveys illustrate well the distinction. Some surveys ask:

(*) "For whom do you expect to vote in the coming election, candidate 0 or candidate 1?"

Others ask:

(**) "For whom would you vote if the election were held today, candidate 0 or candidate 1?"

The former is an intentions question, the latter a forced-choice one.

A person with rational expectations need not give the same response to these two questions. The forced-choice response is the person's decision given the information s available at the time of the survey. The intentions response, on the other hand, is the person's prediction of the decision he will make when he has the information s and z.

The difference between intentions and forced-choice responses shows clearly if we assume that the person maximizes expected utility. (Until now we have imposed no restrictions on the person's decision rule.) Let V(s,z) denote the difference between the expected utilities the respondent associates with candidates 1 and 0, given the information (s,z). Let y be the candidate for whom the respondent actually votes. Let i be the respondent's response to the intentions question (*). Let j be his response to the forced-choice question (**).

Expected utility maximization and rational expectations imply that

$$
\begin{aligned}
(14)\quad y &= 1 \quad \Rightarrow \quad V(s,z) \geq 0 \\
y &= 0 \quad \Rightarrow \quad V(s,z) \leq 0.
\end{aligned}
$$

It follows from this and from (2) that

$$
\begin{aligned}
(15)\quad i &= 1 \quad \Rightarrow \quad P[V(s,z)\geq 0|s] \geq 1/2 \\
i &= 0 \quad \Rightarrow \quad P[V(s,z)\geq 0|s] \leq 1/2.
\end{aligned}
$$

In the forced-choice setting, the respondent maximizes expected utility conditional on the information s available at the time of the survey. Hence

$$
\begin{aligned}
(16)\quad j &= 1 \quad \Rightarrow \quad E[V(s,z)|s] \geq 0 \\
j &= 0 \quad \Rightarrow \quad E[V(s,z)|s] \leq 0.
\end{aligned}
$$

Responses i and j clearly need not be the same. For example, suppose that the pre-election survey is conducted a week before the election. Suppose that candidate 1 will undergo a medical examination the day following the survey. Let z = 1 denote the event that the candidate is found healthy and $z = 0$ that he has a severe illness. Suppose that a person prefers candidate 1 if the candidate is healthy but prefers candidate 0 otherwise; that is, $V(s,z=1) > 0 > V(s,z=0)$. Then this person will respond $i = 1$ if $P(z=1|s) > 1/2$. He will respond $j = 1$ if $P(z=1|s)V(s,z=1) + P(z=0|s)V(s,z=0) > 0$.

The difference between (15) and (16) has a particularly simple interpretation if a regularity condition holds. Suppose that the distribution of $V(s,z)$ conditional on s is continuous and strictly increasing at the point $V(s,z) = 0$. Then (15) is equivalent to

$$(15)' \quad \begin{aligned} i &= 1 \quad \Rightarrow \quad M[V(s,z)|s] \geq 0 \\ i &= 0 \quad \Rightarrow \quad M[V(s,z)|s] \leq 0, \end{aligned}$$

where $M[V(s,z)|s]$ denotes the median of $V(s,z)$ conditional on s. Thus a person with rational expectations will respond differently to intentions and forced-choice questions if the conditional median and expectation of $V(s,z)$ differ in sign.

The foregoing suggests that, given the objective of predicting future behavior, intentions questions are preferable to forced-choice ones. Intentions responses are informative even if the researcher possesses no knowledge of the decision rule determining behavior. Interpretation of forced-choice responses, on the other hand, requires prior knowledge. In the expected utility context, the researcher must know the relationship between the median and expectation of V(s,z). Even given this knowledge, all that a forced-choice question does is to provide the same information about future behavior as does an intentions question.

The conclusion that intentions questions are preferable to forced-choice ones does depend on the maintained assumption that one wishes to predict future behavior. Suppose instead that the objective is to learn about the function V(*,*) and probability distribution $P_z|s$. Then the two types of question have complementary appeal. An intentions (forced-choice) question allows one to learn the sign of the median (expectation) of V(s,z) conditional on s. Thus the two questions are informative about different properties of V(*,*) and $P_z|s$.

6. CONCLUSION

The use of intentions data to predict behavior has been controversial. The analysis presented here suggests that at least some of the controversy is rooted in the flawed premise that discrepancies between intentions and behavior show individuals to be poor predictors of their futures. Discrepancies may simply reflect the dependence of behavior on events not yet realized at the time of the survey. Discrepancies will occur even if responses to intentions questions are the best predictions possible given the available information. The lesson is that researchers should not expect too much from intentions data.

Do intentions data contain useful information on behavior? The answer is certainly positive if individuals respond as would persons with rational expectations. The bounds developed in Section 2 then impose meaningful, informative restrictions on behavior.

On the other hand, it is premature to reach any conclusion on the descriptive validity of the rational-expectations hypothesis. The limited evidence given in Section 3 was mixed. In presenting that evidence I intentionally did not try to explain the several observed violations of the bounds. To understand how people actually respond to survey questions requires serious research, not ex-post rationalizations of particular response patterns.

References

Bierens, H. 1987. "Kernel Estimators of Regression Functions." In T. Bewley, ed., Advances in Econometrics, Vol. I. New York: Cambridge University Press.

Manski, C. 1988. Analog Estimation Methods in Econometrics. London: Chapman and Hall.

Manski, C. and D. Wise. 1983. College Choice in America. Cambridge: Harvard University Press.

O'Connell, M. and C. Rogers. 1983. "Assessing Cohort Birth Expectations Data from the Current Population Survey, 1971-1981." Demography 20: 369-383.

Prakasa Rao, B. L. S. 1983. Nonparametric Functional Estimation. Orlando: Academic Press.

Riccobono, J., L. Henderson, G. Burkheimer, C. Place, and J. Levinsohn. 1981. National Longitudinal Study: Data File Users Manual. Washington, D.C.: National Center for Educational Statistics, U.S. Department of Education.

U.S. Bureau of the Census. 1988. Current Population Reports, Series P-20, No. 427, Fertility of American Women: June, 1987. Washington, D.C.: U.S. Government Printing Office.

Webster's Ninth New Collegiate Dictionary. 1985. Springfield, Ill.: Merriam-Webster.

Westoff, C. and N. Ryder. 1977. "The Predictive Validity of Reproductive Intentions." Demography 14: 431-453.