NONPARAMETRIC SURVEY RESPONSE ERRORS

Rosa L. Matzkin*

Department of Economics

Northwestern University

October 2005

Abstract

We present nonparametric methods to identify and estimate the biases associated with response errors. When applied to survey data, these methods can be used to correct for those biases, analyze how observable characteristics of the respondent and the design of the survey affect the biases, and to design better surveys. We consider cases where the distribution of the true response is known as well as cases where this distribution is unknown. In each case, we allow the response to be influenced by characteristics of the respondent and of the design of the survey. Several cases of statistical dependence between the true response and the observable characteristics are considered.

———————————————————————————————

* The support of the National Institute of Aging for this research is gratefully acknowledged. The work presented in this paper has greatly benefitted from the input of Daniel McFadden and Joseph Winter. I have also benefitted from comments of participants at the joint UC Berkeley/RAND Workshop on Response Errors in Surveys of the Elderly and Internet Interviewing (January 2004) and the Conference in Honor of Daniel McFadden (May 2005).

1. Introduction

Surveys have been extensively used in economics, marketing, sociology, and political science, among other fields. They provide a quick and relatively inexpensive method for gathering data on individuals. Some of this data might be impossible to get through any other way. In surveys, a representative sample of individuals is asked, either verbally or in written form, to respond to a series of questions. These may include questions about factual aspects of the respondent's life, such as age, gender, and marital status; hypothetical questions, such as what the respondent would do in a future situation; or opinions such as approval or disapproval of some government action.

As with any other method, surveys have their own drawbacks. Survey responses are typically plagued by response errors. (Battistin (2003), Bound and Krueger (1991), McFadden, Schwarz and Winter (2004), Philipson (1997, 2001), Poterba and Summers (1986), Schwarz, Hippler, Deutsch, and Strack (1985), Tourangeau, Rips and Rasinski (2000), and the references in Bound, Brown, and Mathiowetz (2001) are some of the works that provide strong evidence for the existence of response errors.) Respondents may have faulty memories, might not interpret a question correctly, or they might be concerned about confidentiality and therefore provide only "socially acceptable" responses, for example. The way in which questions are made and the sequence in which they are made are also well known to affect responses. All these produce biases in the response, which if not dealt with, invalidate the conclusions obtained from survey data.

In this paper, we develop several nonparametric methods to deal with the identification and estimation of survey response errors. The methods will allow one to identify the form of and to measure the noise generated from different sources. Hence, these methods can then be used to (i) predict biases, due to response errors, in new surveys for either the same or a new population of respondents, (ii) "undo" the biases due to response errors, by using the measured errors, and (iii) design surveys in a way that will decrease survey response errors.

One of the methodologies that is commonly used as a first step when analyzing response errors uses descriptive tools to analyze the relationship between responses and a few observable variables. This analysis may look at how the average response to some question varies as the value of some observable characteristics, either of the respondents or of the survey, change. For example, one may ask respondents about their life expectancy and see how the average response changes across different age groups. This "nonparametric, reduced form" analysis is typically used to provide evidence about the existence of some particular effect. It is useful to uncover relationships among a few observable variables and to make simple predictions. However, it can seldom be used to predict what will happen in a new situation, such as when the values of other variables change. A more structural model is typically needed if one wants to measure the relationships, incorporate unobservable variables, model the interaction among different errors, and analyze and measure the effect of different characteristics of the survey design.

Another commonly used structural approach proceeds by specifying functions and distributions as known, up to a few parameters. For example, in a yes or no answer, this approach would proceed by first specifying that a particular individual will answer yes if the value of an unobservable variable is above some threshold; where the value of this variable is a linear function of some observable and unobservable characteristics of the respondent. Using data on each individual's responses and observable characteristics, this method provides numbers for the coefficients of the linear function, which could then be used to analyze, predict, and correct errors in the response of individuals. This "parametric, structural form" analysis has been typically used to measure the effect of a variable in a certain response, while controlling for other variables that may also affect this response, and to uncover the distributions of relevant unobservable variables (See, for example, Hurd, McFadden, et al. (1998)).

The methodology that we present in this paper provides a "nonparametric, structural form" analysis. Hence, it is located between the two common but very different approaches described above. The method is structural, because it allows one to estimate underlying functions and distributions of key unobservable variables, in different stages of the response process. The method is nonparametric, because it does not require specifying a-priory parametric structures for the underlying functions and distributions. These new methods can be used with minimal or full blown structural models, or anywhere in between.

As an example of how one can use the new methods to add a minimal amount of structure to a reduced form model, suppose that one is interested in understanding the variation in the response of individuals that are otherwise equal in their relevant observable characteristics. For example, suppose that individuals are asked their perceived probability of an end-of-life health hazard, such as needing nursing home care. Health status and family status will be typically the observable characteristics that one may plot this answer against. However, another important variable, which is unobservable but can explain the variation in the response of individuals within a common health and family status, is the attitude of the respondent towards living in a nursing home. Being able to identify and estimate the distribution of attitudes towards living in a nursing home, and the variables that affect this taste distribution, is very important to predict future demand for nursing homes and to measure the well-being of nursing home users. Moreover, this distribution of attitudes towards living in a nursing home will also influence the variation in the response to other related questions, such as "Have you purchased insurance for nursing home care?" The response to the latter question may depend on a larger set of observable and unobservable characteristics than the former question, which may include the income and the unobservable attitude toward risk of the respondent. Identifying the distribution of attitudes towards living in a nursing home, from the response to the first question, will make it feasible to identify the distribution of the unobservable attitude towards risk, from the responses to the second question. Analyzing the latter distribution will have important implications to predict demand for various types of insurance, and its identification will allow to pursue the identification of other important unobservable variables as well as other further studies in which knowing this distribution is important.

This nonparametric structural approach provides many benefits. First, it provides a bridge among the two different methods described above, which have been used to analyze behavior in survey response. Second, it provides more "trusted" predictions than the parametric analysis, because its conclusions do not depend on ad-hoc parametric specifications for the underlying functions and distributions. Third, it provides a method to test particular parametric assumptions, by evaluating how close the nonparametric estimates are from the parametric ones. Fourth, it allows to infer all types of shapes for the structural anomalies, because the underlying functions will be estimated without imposing on them particular shapes. And, fifth, it allows one to infer the unobserved heterogeneity across otherwise observable equal individuals, which has implications towards their heterogenous responses in other questions and towards predicted behavior by these individuals.

Rather than looking into identification of response functions, as we do in this paper, one could consider estimating bounds on the functions of interest. (See Horowitz and Manski (1995), Molinari (2005), and the other references on the survey article by Manski (2005).) Also, the models that we consider can be interpreted as particular nonparametric measurement error models. (See Hsiao (1989), Chesher (1991, 1998), Hausman, Newey, Ichimura, and Powell (1991), Fan (1991), Fan and Truong (1993), the survey in Carroll, Ruppert, Stefanski (1995), Hausman, Newey and Powell (1995), Wang and Hsiao (1995), Lewbel (1998), Li and Vuong (1998), Hsiao and Wang (2000), Newey (2001), Li (2002), Wang (2002), Schennach (2004, 2005), and Chen, Hong and Tamer (2005) for some nonlinear and nonparametric methods.) Using nonparametric regressions, Bollinger (1998) analyses measurement error in the Current Population Survey.

The outline of the paper is as follows. In the next section, we describe the model and describe the notation. Section 3 deals with cases where the distribution of an unobservable variable of interest is known and this variable is distributed independently of the other variables that affect a response. Section 4 deals with cases where the distribution of the unobservable variable of interest is unknown. We deal in this section with three cases. In the first case, the unobservable variable of interest is distributed independently of other variables that affect a response. We show how the response function and response errors are identified, and describe how these can be estimated nonparametrically. In the second case, the unobservable variable is not distributed independently of the othe variables that affect response. We present several cases where augmenting the data in certain ways, one can obtain a conditional independence property between the unobservable variable of interest and those other variables. This property provides ways of identifying the response functions and the distributions of the unobservable variables. The third case is as the second one, in that the unobservable variable of interest is not distributed independently of other variables that affect response. But, unlike the second case, augmenting the data does not guarantee a conditional independence property. We show how, nevertheless, one can identify the response function, using an observable instrument. Section 5 provides conclusions, and describes extensions of the methods presented in the main sections. All the proofs are presented in the Appendix.

2. The model

Consider a situation where one is interested in inferring the true value of a latent variable, Y^*, such as the life expectancy a respondent thinks he has. The individual may be asked directly to state Y^*. He may be asked indirectly about Y^*, like asking whether he thinks he will live more or less than the average in some population. He may be asked about other variables that would affect his perception of his life expectancy, like the mortality experience of his parents. Or, he may be asked about decisions that would be affected by what he thought his life expectancy is, such as his saving or whether he purchased life-insurance. Any of these formats is prompt to response errors. In the next sections, we will present several methods to estimate these errors nonparametrically, under different situations. We will denote by W a vector of observable variables characterizing features of the design of the questions. These could be any of the features of the survey questions that are known to affect responses, such as any numbers in the survey indicating average responses in a population, or whether the questions were done by internet, phone, or mail. We will denote by X a vector of observable variables of characteristics of the respondent. In Section 3 we will consider situations where the distribution of Y^* is known and Y^* is distributed independently of the variables X and

W. In Section 4, we will deal with situations where the distribution of Y^* is unknown, and Y^* is not distributed independently of (X, W).

3. Known distribution of the true value Y^*

In many situations, one may be able to know the distribution of some variable of interest but not be able to know the particular value of such variable for any particular individual. For example, one may know the distribution of some objective measure, Y^*, of health status in a population. Suppose that individuals in this population are asked their health status. Let Y denote their response. Assume that the relationship between Y and Y^* is given by

$$(3.1) \qquad Y = m\left(Y^*, X, W\right)$$

where either X or W may possess no coordinates and (X, W) is such that Y^* is independent of (X, W). An example of such a situation is where X contains no coordinates and W denotes some number in the survey which may indicate to the respondent the average response, and which varies randomly across surveys. Let $F_{Y|X,W}$ denote the joint distribution of the observable variables (Y, X, W). Let F_{Y^*} denote the distribution of Y^*. The following result follows from Matzkin (2003):

Theorem 1: *Suppose that Y^* is distributed independently of (X, W) with an everywhere positive, known density. Suppose also that for each (X, W), the function m is strictly* *increasing in Y^*. Then, for all y^*, x, w*

$$(3.2) \qquad m\left(y^*, x, w\right) = F^{-1}_{Y|X=x,W=w}\left(F_{Y^*}(y^*)\right)$$

Hence, the function m is identified nonparametrically from the joint distribution of (Y, X, W).

Equivalently, we can state that, under the assumptions of Theorem 1, an individual that answers $Y = y$ when faced with $(X, W) = (x, w)$ has a value of the latent variable Y^* equal to

$$y^* = F^{-1}_{Y^*}(F_{Y|X=x,W=w}\left(y\right))$$

Replacing $F_{Y|X,W}$ (and F_{Y^*}) by a nonparametric estimator, one can obtain from the above equations estimators for the response function, $m\left(y^*, x, w\right)$, and for the value of the latent variable of any individual, given his response, y.

The assumptions in the above model are such that there is a 1-1 relationship between the true value, Y^*, of the response, and the response, Y. The distributions of Y^* and of Y conditional on (X, W) allow one to determine uniquely such a 1-1 relationship. From this, one can map any response with its true corresponding value, without any error. In some cases, this situation may be unrealistic. The same value of the true latent variable Y^* may generate different responses even from individuals that possess the same observable characteristics, X, and are asked the same question, characterized by W. Moreover, in many cases, such additional response error, η, is not independent of one or more of the coordinates of X or W. As an example, suppose that individuals are asked to respond to the question: "What was the dollar amount of your last phone bill?" Let Y^* denote the true amount, which is verifiable for some individuals. The variable η denotes the recall error, which could depend on the time since the individual last saw the phone bill. The dependence of η on some coordinates, X_1, W_1 of X and W may be modeled by specifying that

$$(3.3) \qquad \eta = v\left(X_1, W_1, \delta\right)$$

where v is an unknown function that is strictly increasing in δ, and δ is an unobservable random term that is distributed independently of (X, W, Y^*). W_1 may denote an indicator for whether or not the individual is asked to look at the phone bill before answering. (Either X_1 or W_1 may contain no coordinates.) Partition X and W by $X = (X_1, X_2)$ and $W = (W_1, W_2)$. Suppose that the model that determines the response Y as a function of Y^*, η, X and W is given by

$$(3.4) \qquad Y = m\left(Y^* + \eta, X_2, W_2\right)$$

This can be seen as a generalization of the standard measurement error problem, where Y^* is observed with error η. The measurement, Y, does not need to be linear in $Y^* + \eta$, and this mapping can depend on some other variables, (X_2, W_2). The following theorem establishes the identification of the functions v and m and of the distributions of δ and η, under some assumptions.

Theorem 2: *Suppose that (Y^*, δ) is distributed independently of (X, W) with an everywhere positive density, δ is distributed independently of (X, W, Y^*) with an everywhere positive un-*

known density, the function m *is strictly increasing in* $Y^* + \eta$, *the function* v *is strictly increasing in* δ, *and the distribution of* Y^* *is known and its characteristic function is everywhere different zero. Restrict the function* v *to satisfy at one point* $(\overline{x}_1, \overline{w}_1)$ *of* (X_1, W_1) *the condition:* $v(\overline{x}_1, \overline{w}_1, \delta) = 0$ *and at another point* $(\widetilde{x}_1, \widetilde{w}_1)$ *of* (X_1, W_1) *the condition:* $v(\widetilde{x}_1, \widetilde{w}_1, \delta) = \delta$. *Then, the function* m, *the function* v, *and the distributions of* δ *and of* η *conditional on* (X_1, W_1) *are identified nonparametrically from the joint distribution of* (Y, X, W).

In the phone bill example, where η depends either on X_1, which denotes time since the individual last saw the phone bill or on W_1, which equals 0 if the individual is asked to look at the phone bill and it equals 1 otherwise, one would naturally let $\widetilde{x}_1 = 0$ and $\widetilde{w}_1 = 0$. If the function v were specified as: $v(x_1, w_1, \delta) = \delta x_1 + \delta w_1$, then the restrictions on the function v would be satisfied for $\widetilde{x}_1 = \widetilde{w}_1 = 0$, and for $(\overline{x}_1, \overline{w}_1) = (1, 0)$ or $(\overline{x}_1, \overline{w}_1) = (0, 1)$.

Theorem 2 establishes the nonparametric identification of the response error, $v(x_1, w_1, \delta)$, and the response function m, when the distribution of the variable of interest, Y^* is known. Since the proof is constructive, one can use the proof to derive nonparametric estimators for these functions. Using similar reasonings to those used in the proof of Theorem 2, one can establish the identification of other particular structures. Suppose for example that instead of the model described in (2.3)-(2.4), we specify that for some unknown functions m and δ

$$(3.5) \qquad Y = m(Y^*, X_2, W_2)\ \eta$$

where η is, as above, an unobservable random term whose distribution depends on (X_1, W_1), so that

$$(3.6)\qquad \eta = v\left(X_1, W_1, \delta\right)$$

Then, one can establish the following

Theorem 3: *Suppose that (Y^*, δ) is distributed independently of (X, W) with an everywhere positive density, δ is distributed independently of (X, W, Y^*) with an everywhere positive unknown density, the function m is strictly increasing in Y^*, the function v is strictly increasing in δ, and the distribution of Y^* is known and its characteristic function is everywhere different from 0. Restrict the function v to satisfy at one point $(\overline{x}_1, \overline{w}_1)$ of (X_1, W_1) the condition: $v\left(\overline{x}_1, \overline{w}_1, \delta\right) = 1$ and at another point, $(\widetilde{x}_1, \widetilde{w}_1)$ of (X_1, W_1) the condition: $v\left(\widetilde{x}_1, \widetilde{w}_1, \delta\right) = \delta$ Then, the function m, the function v, and the distribution of δ and of η conditional on (X_1, W_1) are identified nonparametrically from the joint distribution of (Y, X, W).*

4. Unknown distribution of the true value Y^*

The analysis in Section 3 rested on the assumption that the distribution of Y^* was known or could be estimated. In some cases, this might not be a feasible situation. To be able to identify the response function m and the response error v nonparametrically will require making some additional assumptions. These may take the form of imposing additional shape restrictions on the functions, or of making a normalization either on the function m or on the distribution of Y^*.

4.1. Y^* independent of the observable variables

In some cases, we might be faced with a situation where one component of the response is recalled without error, while the other is recalled with error. For example, if we ask an individual how much he has spent in restaurants in the last month, he might be able to recall the part of this expense that he regularly incurs in, but he will probably not recall perfectly the irregular expenses. Let X_1 denote the regular expenses, which are perfectly recalled and let Y^* denote the irregular expenses. Let η denote the recall error and let Y denote the response. Suppose that we observe or believe the number given for the regular expenses, X_1. Then, we can specify the model as

$$(4.1) \quad Y = m\left(Y^* + \eta + X_1, X_2, W_2\right), \text{ with}$$

$$(4.2) \quad \eta = v\left(W_1, \delta\right)$$

The variable W_1 may denote, for example, the time frame that the question refers to. We will assume that X_1 has support equal to R and that Y^* is distributed independently of (X, W). The following theorem is proved in the Appendix.

Theorem 4: *Suppose that (Y^*, δ) is distributed independently of (X, W) with an everywhere positive unknown density, δ is distributed independently of (X, W, Y^*) with an everywhere positive unknown density, the function m is strictly increasing in $Y^* + \eta$, the function v is strictly increasing in δ, and the characteristic function of Y^* is everywhere different from 0. Restrict the function v to satisfy at one point $\overline{w}_1$ of W_1 the condition: $v\left(\overline{w}_1, \delta\right) = 0$ and at another point $\widetilde{w}_1$ of W_1 the condition: $v\left(\widetilde{w}_1, \delta\right) = \delta$. Restrict the function m to* *satisfy at a point* $\left(\bar{t}, \overline{x}_2, \overline{w}_2\right)$ *the restriction* $m\left(\bar{t}, \overline{x}_2, \overline{w}_2\right) = \alpha$*. Suppose that the distribution of* Y^* *is unknown. Then, the function* m*, the function* v*, and the distribution of* δ *and of* η *conditional on* (X_1, W_1) *are identified nonparametrically from the joint distribution of* (Y, X, W).

4.2. Y^* conditionally independent of the explanatory variables.

A common method used for inferring Y^*, when Y^* is unknown, is to ask a question about an indicator for Y^*. For example, one may ask a question about the amount of life insurance an individual has purchased, when Y^* is expected mortality. In such cases, the answer will depend not only on Y^*, but also on some observable variables, such as savings. The assumption that Y^* is distributed independently of savings would not be a good assumption. Suppose then that the model is

$$(4.3) \qquad Y = m\left(Y^*, X\right)$$

and that Y^* and X are not be independently distributed. One possibility to deal with this situation is to ask another question, whose answer is an exogenous perturbation of X, in the sense that for some unknown function s and some unobservable ω

$$(4.4) \qquad X = s\left(\widetilde{X}, \omega\right)$$

where ω is such that ω is independent of Y^* conditionally on $\widetilde{X}$. For example, if X is savings at the present time, $\widetilde{X}$ might be savings in some recent time. The unobservable ω could then represent unobservables that affect X but are independent of Y^* given $\widetilde{X}$, such as investment returns that were unexpected at the time when savings were $\widetilde{X}$. If ω can be assumed to be independent of Y^* conditional on at least one value $\overline{\widetilde{x}}$ of $\widetilde{X}$, then, the results in Matzkin (2004) can be directly applied to this situation to identify and nonparamerically estimate the function m and the joint distribution of (Y^*, X).

If we can find an observable Z such that Z is independent of (Y^*, ω) and

$$X = s(Z, \omega)$$

for some unknown function s that is strictly increasing increasing in ω, then the results in Imbens and Newey (2003) and in Chesher (2003) could be used to identify m and the distribution of (Y^*, X). In the above example, where X denotes savings, Z could denote the price of housing.

Alternatively, we may have information about some of the determinants of Y^*. Consider, for example, the model where

$$Y = m(Y^* + \eta, X_2, W_2), \text{ and}$$

$$\eta = v(X_1, W_1, \delta)$$

Suppose that for some unknown function s, some vector, Z, of observable variables, and an unobservable variable, ξ,

$$Y^* = s(Z, \xi)$$

This model may describe, for example, a situation where the respondent is asked the cost of the drugs that he has used in a number of specified last months. The respondent may have insurance to pay for some of those drugs, in which case he might be able to only poorly estimate the costs of those drugs. Moreover, the larger the period over which he has to calculate his expenses, the more likely he will forget about over-the-counter drugs that he had bought. The true expense, Y^*, depends on observable determinants, such as the health conditions the respondent has been diagnosed with, and other unobservable determinants. In this example, Z may denote those health conditions, X_1 may denote the type of prescription drug insurance that the respondent has, W_1 may denote the number of months the respondent is asked to calculate his expenses for, and (X_2, W_2) may denote other characteristics of the respondent and of the question design that affect the individual's response.

Under assumptions similar to those made in Theorem 2, and some additional assumptions on s and ξ, one can show that the functions m, v, and s are identified. The later set of assumptions may be imposed on the function s, similarly to those imposed in Theorem 2 on the function v and δ, or on the distribution of ξ. We consider the former.

Theorem 5: *Suppose that (ξ, δ) is distributed independently of (X, W, Z) with an everywhere positive unknown density, δ is distributed independently of (X, W, Z, ξ) with an everywhere positive unknown density, the function m is strictly increasing in $Y^* + \eta$, the function v is strictly increasing in δ, the function s is strictly increasing in ξ, and the characteristic function of ξ is everywhere different from 0. Restrict the function v to satisfy at one point $(\overline{x}_1, \overline{w}_1)$ of (X_1, W_1) the condition: $v(\overline{x}_1, \overline{w}_1, \delta) = 0$ and at another point $(\widetilde{x}_1, \widetilde{w}_1)$ of (X_1, W_1) the condition: $v(\widetilde{x}_1, \widetilde{w}_1, \delta) = \delta$. Restrict the function s to satisfy at one point $\overline{z}$ of Z the condition: $s(\overline{z}, \xi) = \xi$. Restrict the function m to satisfy at one point $(\overline{x}_2, \overline{w}_2)$*

of (X_2, W_2) *the condition:* $m(t, \overline{x}_2, \overline{w}_2) = t$. *Then, the functions* m, v, *and* s *as well as the distributions of* δ, η, ξ, *and* Y^* *conditional on* (X, W, Z) *are identified nonparametrically from the joint distribution of* (Y, X, W, Z).

In the above example, in which an individual is asked to respond about his drug expenses, $\widetilde{z}$ may denote the situation where the respondent has no health condition that requires medication, and hence the drug costs are composed of sporadic costs for over-the-counter medications, the function v is the recall error function, which may equal 0 when the respondent is asked to recall expenses incurred during a short period ($\overline{w}_1 = 0$) and when his insurance does not pay for drugs or pays a given fixed percentage ($\overline{x}_1 = 0$).

4.3. Y^* not conditionally independent of the explanatory variables.

Not in all cases it is reasonable to assume that Y^* is conditionally independent of some variables. In many cases, we might encounter a simultaneity situation, where some of the observable variables, X, that determine the response, Y, together with the true value, Y^*, are themselves partly determined by Y. As an example, suppose that Y is the number of chronic health problems that the respondent thinks he has, Y^* is the true number of chronic health problems that he has, and X is the number of times in a year that the respondent goes to the doctor. Disregarding design variables, W, the model for Y could then be described as

$$(4.3) \quad Y = m(Y^*, X)$$

Clearly, the number of times the respondent goes to the doctor is affected by the perceived number of chronic conditions. Hence, it is not only the case that Y is a function of X, but also X is a function of Y. We can deal with this situation by looking for an observable instrument, Z that is independent of Y^*. For example, Z might be the number of doctors in a nearby distance from where the respondent lives. The instrument Z is independent of Y^* and X can be expressed as a function of Z. Since such Z allows for a simultaneity between X and Y, we model the relationship between X and Z by

$$(4.4) \qquad X = s\left(Z, Y, \omega\right)$$

for some unobservable ω. We assume that Z is distributed independently of $\left(Y^*, \omega\right)$, that m is strictly increasing in Y^*, and that s is strictly increasing in ω. Hence, letting r_1 denote the inverse of m with respect to Y^*, and letting m_2 denote the inverse of s with respect to ω, it follows that

$$\begin{aligned} Y^* &= r_1\left(Y, X\right) \\ \omega &= r_2\left(Y, X, Z\right) \end{aligned}$$

The nonparametric identification of the functions r_1 and r_2 as well as of the distribution of (Y^*, ω) can be analyzed by specializing the results in Matzkin (2005a). Consider a set of functions $(\widetilde{r}_1, \widetilde{r}_2)$ to which (r_1, r_2) belong. We will make use of the following assumption.

Assumption A: *All the observable and unobservable variables have differentiable and everywhere positive densities. The functions $\widetilde{r}_1$ and $\widetilde{r}_2$ are continuously differentiable. For* *all y, x, z, the determinant of the matrix*

$$\begin{pmatrix} \frac{\partial \widetilde{r}_1(y,x)}{\partial y} & \frac{\partial \widetilde{r}_1(y,x)}{\partial x} \\ \frac{\partial \widetilde{r}_2(y,x,z)}{\partial y} & \frac{\partial \widetilde{r}_2(y,x,z)}{\partial x} \end{pmatrix}$$

is 1. For all y, x, z, $\partial \widetilde{r}_2(y,x,z)/\partial z \neq 0$. For all $\widetilde{r}_1 \neq r_1$, there exists (y, x, z) such that

$$\frac{\frac{\partial r_1(y,x)}{\partial x}}{\frac{\partial r_1(y,x)}{\partial y}} \neq \frac{\frac{\partial \widetilde{r}_1(y,x)}{\partial x}}{\frac{\partial \widetilde{r}_1(y,x)}{\partial y}},$$

$\partial f_{Y^*,\omega}\left(r_1(y,x), r_2\left(y,x,z\right)\right)/\partial Y^* \neq 0$ *and* $\partial f_{Y^*,\omega}\left(r_1(y,x), r_2\left(y,x,z\right)\right)/\partial \omega = 0.$

The following theorem establishes the identification of the function r. Estimation of these functions can be performed along the lines in Matzkin (2005b)

Theorem 6: *Suppose that $(\widetilde{r}_1, \widetilde{r}_2)$ and (r_1, r_2) satisfy Assumption A. Suppose that $\widetilde{r}_1 \neq r_1$. Then, $(\widetilde{r}_1, \widetilde{r}_2)$ and any distribution of (Y^*, ω) that is independent of Z cannot generate the same distribution of the observable variables as (r_1, r_2) and $F_{Y^*,\omega}$ do.*

5. Extensions

The results that have been presented in the previous sections can be extended to analyze models with nested response errors and models where the responses are discrete.

Answering a survey question involves the stage of comprehension of the question, the stage of retrieving and assembling relevant information, the stage of filtering the information, and the stage of actually responding to the question with an answer which may or may not be the one that the respondent have come up with prior to responding (see Tourangeau, Rips, and Rasinski (2000)). Each of these stages adds a level of noise to the response, which will typically be different across respondents. Variation in the noise may depend on observable characteristics, but, it will typically also depend on unobservable characteristics. The design of the survey and of the particular question being asked add additional layers of possible noise, which interact and affect the magnitudes of the processing noise. A model that allows for all these errors at the different stages may take, for example, the form

$$Y = m\left(s\left(\left(v\left(\delta_3, X_3, W_3\right) + \delta_2\right), X_2, W_2\right) + \delta_1, X_1, W_1\right)$$

or the form

$$Y = m\left(s\left(X_2, W_2, \delta_2\right) + v\left(X_3, W_3, \delta_3\right) + \delta_1, X_1, W_1\right).$$

The identification and estimation of the unknown functions and distributions in these models can be achieved by following a similar analysis to that described in the previous sections.

Discrete responses are also very common. Some questions ask for a 0 or 1 answer. Some respondents prefer not to answer some questions. The identification and estimation of such models can be performed also by extending the above results. Consider for example a situation where an individual is asked whether he thinks that the probability that he will ever need a nursing home is above or below .5. The individual may first decide whether to answer the question. Let $R = 1$ if he answers and $R = 0$ otherwise. The individual may decide not to answer the question if he does not have any good information about the matter, or he has not thought about it in the past. Some of the variables affecting this behavior may be observable and some may be unobservable. Suppose, in particular, that R is an indicator for the event $X_2 + v(X_1, W, \eta) \geq \varepsilon_R$, where X_1, X_2, and W are observable, η and ε_R are unobservable, X_2 is distributed independently of $(X_1, W, \eta, \varepsilon_R)$, ε_R is distributed independently of (X_1, X_2, W, η), and the value of v is known at one point. (This setup generalizes a model of discrete response to survey treatments X_1 analyzed by McFadden (1994).) Using the results in Matzkin (1993, 1994), one can show that from the observable distribution of (R, X_1, X_2, W) one can infer the distribution of the latent variable $V = v(X_1, W, \eta)$ conditional on (X_1, W). This brings one back to the situation where a response V is observable given observable characteristics, (X_1, W). The estimation methods described above can be used then to identify the function V and the distribution of η, conditional and unconditional on (X_1, W_1).

Next, let $Z = 1$ if the response is that the probability is above .5; $Z = 0$ otherwise. Suppose that $Z = 1$ if $\widetilde{X}_2 + r(\widetilde{X}_1, \widetilde{W}, \eta) \geq \varepsilon_Z$, where the assumptions on the observable variables, $\widetilde{X}_1$, $\widetilde{X}_2$, and $\widetilde{W}$, the unobservable variables, η and ε_Z, and the function r are similar to those described above in the model that determines R. Since in this case Z is observed only for the individuals for which $R = 1$, one cannot use the same analysis as above. It is first necessary to identify the joint distribution of $(\varepsilon_1, \varepsilon_2)$. This can be done using the expression

$$\Pr(R = 1, Z = 1 | X, W, \widetilde{X}, \widetilde{W}, \eta) = F_{\varepsilon_R, \varepsilon_Z}\left(X_2 + v(X_1, W, \eta), \widetilde{X}_2 + r(\widetilde{X}_1, \widetilde{W}, \eta)\right)$$

with enough independence, and the fact that the function v and the distribution of η are identified in the first step. These together with a normalization such as $r(\widetilde{x}_1^*, \widetilde{w}^*, \eta)$ constant for all η, guarantee identification of all the elements in the model. (See Briesch, Chintagunta, and Matzkin (2005) and Matzkin (2005a, 2005b).)

6. Appendix

Proof of Theorem 1: By independence between Y^* and (X,W) and strict monotonicity, it follows that for all (x,w)

$$\begin{aligned} F_{Y^*}(y^*) &= \Pr(Y^* \leq y^*) = \Pr(Y^* \leq y^* | X=x, W=w) \\ &= \Pr(m(Y^*,X,W) \leq m(y^*,x,w) \,| X=x, W=w) \\ &= F_{Y|(X,W)=(x,w)}(m(y^*,x,w)) \end{aligned}$$

Since the assumptions imply that $F_{Y|(X,W)=(x,w)}$ is strictly increasing, it follows that

$$m(y^*,x,w) = F^{-1}_{Y|X=x,W=w}(F_{Y^*}(y^*))$$

Hence, m is identified

Proof of Theorem 2: Since (Y^*,δ) is distributed independently of (X,W), Y^* is distributed independently of (X_2,W_2) conditional on any values of (X_1,W_1), $Y^*+\delta$ is distributed independently of (X_2,W_2) conditional on any values of (X_1,W_1), and $Y^*+\eta = Y^*+v(X_1,W_1,\delta)$ is distributed independently of (X_2,W_2) conditional on any values of (X_1,W_1). Moreover, when $(X_1,W_1)=(\overline{x}_1,\overline{w}_1)$, $\eta = v(\overline{x}_1,\overline{w}_1,\delta)=0$ and $Y^*+\eta=Y^*$; and when $(X_1, W_1) = (\widetilde{x}_1, \widetilde{w}_1)$, $\eta = v(\widetilde{x}_1, \widetilde{w}_1, \delta) = \delta$ and $Y^* + \eta = Y^* + \delta$. Hence, for any (x_2, w_2)

$$
\begin{aligned}
&(T2.1) \quad F_{Y^*}(y^*) \\
&= F_{Y^*+\eta}(y^* | (X_1, W_1) = (\overline{x}_1, \overline{w}_1)) \\
&= F_{Y^*+\eta}(y^* | (X_1, W_1) = (\overline{x}_1, \overline{w}_1), (X_2, W_2) = (x_2, w_2)) \\
&= \Pr(Y^* + \eta \leq y^* | (X_1, W_1) = (\overline{x}_1, \overline{w}_1), (X_2, W_2) = (x_2, w_2)) \\
&= \Pr(m(Y^* + \eta, X_2, W_2) \leq m(y^*, x_2, w_2) | (X_1, W_1) = (\overline{x}_1, \overline{w}_1), (X_2, W_2) = (x_2, w_2)) \\
&= F_{Y|X=(\overline{x}_1, x_2), W=(\overline{w}_1, w_2)}(m(y^*, x_2, w_2))
\end{aligned}
$$

Since our assumptions imply that $F_{Y|X=(\overline{x}_1, x_2), W=(\overline{w}_1, w_2)}$ is invertible, this implies that for any t and any (x_2, w_2)

$$(T2.2) \quad m(t, x_2, w_2) = F^{-1}_{Y|X=(\overline{x}_1, x_2), W=(\overline{w}_1, w_2)}(F_{Y^*}(t))$$

Hence, m is identified nonparametrically.

Using a similar reasoning as above, we get that for any (x_2, w_2)

$$
\begin{aligned}
&(T2.3) \quad F_{Y^*+\delta}(y^* + \delta) \\
&= F_{Y^*+\eta}(y^* + \delta | (X_1, W_1) = (\widetilde{x}_1, \widetilde{w}_1)) \\
&= F_{Y^*+\eta}(y^* + \delta | (X_1, W_1) = (\widetilde{x}_1, \widetilde{w}_1), (X_2, W_2) = (x_2, w_2)) \\
&= \Pr(Y^* + \eta \leq y^* + \delta | (X_1, W_1) = (\widetilde{x}_1, \widetilde{w}_1), (X_2, W_2) = (x_2, w_2)) \\
&= \Pr(m(Y^* + \eta, X_2, W_2) \leq m(y^* + \delta, x_2, w_2) | (X_1, W_1) = (\widetilde{x}_1, \widetilde{w}_1), (X_2, W_2) = (x_2, w_2)) \\
&= F_{Y|X=(\widetilde{x}_1, x_2), W=(\widetilde{w}_1, w_2)}(m(y^* + \delta, x_2, w_2))
\end{aligned}
$$

Using (T2.1), this implies that

$$F_{Y^*+\delta}(y^*+\delta) = F_{Y|X=(\widetilde{x}_1,x_2),W=(\widetilde{w}_1,w_2)}\left(F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2)}\left(F_{Y^*}(y^*+\delta)\right)\right)$$

Hence, for any t

$$(T2.4) \qquad F_{Y^*+\delta}(t) = F_{Y|X=(\widetilde{x}_1,x_2),W=(\widetilde{w}_1,w_2)}\left(F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2)}\left(F_{Y^*}(t)\right)\right)$$

This implies that the distribution of $Y^*+\delta$ is identified. Since, by assumption, the distribution of Y^* is known, one can obtain the distribution of δ by deconvolution. Hence, the distribution of δ is identified.

Next, we derive an expression for the distribution of $Y^*+\eta$ conditional on any (X_1,W_1). Similarly to above,

$$\begin{aligned}
&(T2.5) \qquad F_{Y^*+\eta|(X_1,W_1)=(x_1,w_1)}(y^*+\eta) \\
&= F_{Y^*+\eta}\left(y^*+\eta|\left(X_1,W_1\right)=\left(x_1,w_1\right),\left(X_2,W_2\right)=\left(x_2,w_2\right)\right) \\
&= \Pr\left(Y^*+\eta \le y^*+\eta|\left(X_1,W_1\right)=\left(x_1,w_1\right),\left(X_2,W_2\right)=\left(x_2,w_2\right)\right) \\
&= \Pr\left(m\left(Y^*+\eta,X_2,W_2\right) \le m\left(y^*+\eta,x_2,w_2\right)|\left(X_1,W_1\right)=\left(x_1,w_1\right),\left(X_2,W_2\right)=\left(x_2,w_2\right)\right) \\
&= F_{Y|X=(x_1,x_2),W=(w_1,w_2)}\left(m\left(y^*+\eta,x_2,w_2\right)\right) \\
&= F_{Y|X=(x_1,x_2),W=(w_1,w_2)}\left(F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2)}\left(F_{Y^*}(y^*+\eta)\right)\right)
\end{aligned}$$

where the last equality follows by (T2.2). Hence, for any t and any (x_1,w_1),

$$(T2.6) \qquad F_{Y^*+\eta|(X_1,W_1)=(x_1,w_1)}(t) = F_{Y|X=(x_1,x_2),W=(w_1,w_2)}\left(F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2)}\left(F_{Y^*}(t)\right)\right)$$

This implies that the distribution of $Y^* + \eta$ conditional on $(X_1, W_1) = (x_1, w_1)$ is identified. Since, by our assumptions, Y^* is distributed independently of η conditional on (X_1, W_1), Y^* is distributed independently of (X_1, W_1), and the distribution of Y^* is known, we can get from the distribution of $Y^* + \eta$ conditional on (X_1, W_1) the distribution of η conditional on (X_1, W_1), by deconvolution. Hence, the distribution of η conditional on (X_1, W_1) is identified.

Last, to show that the function v is identified, we use the strictly monotonicity of v in δ and the independence between δ and (X_1, W_1), to establish as in the proof of Theorem 1 that for any t and any (x_1, w_1)

$$F_\delta(t) = F_{\eta|(X_1,W_1)=(x_1,w_1)}\left(v(x_1, w_1, t)\right)$$

This implies that

$$(T2.7) \qquad v(x_1, w_1, t) = F^{-1}_{\eta|(X_1,W_1)=(x_1,w_1)}\left(F_\delta(t)\right).$$

Since F_δ and $F_{\eta|(X_1,W_1)=(x_1,w_1)}$ are identified, v is identified. This completes the proof.

Proof of Theorem 3: Denote by Y_1 the value of $m(Y^*, X_2, W_2)$. Then, $Y = Y_1\eta$, and by our assumptions, Y_1 and η are independently distributed conditional on (X, W). Moreover, when $(X_1, W_1) = (\overline{x}_1, \overline{w}_1)$, $Y = Y_1$. Hence, for any y, x_2, w_2

$$\begin{aligned}
& F_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,W)}(y) \\
= \; & F_{Y_1|X=(\overline{x}_1,x_2),W=(\overline{w}_1,W)}(y) \\
= \; & \Pr\left(Y_1 \leq y | X = (\overline{x}_1, x_2), W = (\overline{w}_1, w_2)\right) \\
= \; & \Pr\left(m\left(Y^*, X_2, W_2\right) \leq y | X = (\overline{x}_1, x_2), W = (\overline{w}_1, w_2)\right) \\
= \; & \Pr\left(Y^* \leq m^{-1}\left(y, x_2, w_2\right) | X = (\overline{x}_1, x_2), W = (\overline{w}_1, w_2)\right) \\
= \; & F_{Y^*}\left(m^{-1}\left(y, x_2, w_2\right)\right)
\end{aligned}$$

where m^{-1} denotes the inverse of m with respect to its first coordinate, and where the last equality follows because Y^* is distributed independently of (X, W). It follows that for any t

$$m\left(t, x_2, w_2\right) = F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,W)}\left(F_{Y^*}(t)\right)$$

Hence, m is identified.

Denote $\ln(Y)$ by $\widetilde{Y}$, $\ln(\eta)$ by $\widetilde{\eta}$, and $\ln\left(Y_1\right)$ by $\widetilde{Y}_1$. Then, $\widetilde{Y} = \widetilde{Y}_1 + \widetilde{\eta}$, the distribution of $\widetilde{Y}$ conditional on (X, W) is known, and, for any (x, w), the distribution of $\widetilde{Y}_1$ conditional on (X, W) is also known since for any t

$$\begin{aligned}
& \Pr\left(\widetilde{Y}_1 \leq t | X = x, W = w\right) \\
= \; & \Pr\left(\widetilde{Y}_1 \leq t | X = (\overline{x}_1, x_2), W = (\overline{w}_1, w_2)\right) \\
= \; & \Pr\left(\widetilde{Y} \leq t | X = (\overline{x}_1, x_2), W = (\overline{w}_1, w_2)\right) \\
= \; & F_{\widetilde{Y}|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2)}(t)
\end{aligned}$$

Moreover, by our assumptions, for any (x, w), $\widetilde{Y}_1$ and η are distributed independently, conditional on $(X, W) = (x, w)$. Hence, from the distribution of $\widetilde{Y}$ conditional on $(X, W) = (x, w)$ and the distribution of $\widetilde{Y}_1$ conditional on $(X, W) = (x, w)$, one can obtain the distribution of $\widetilde{\eta}$ conditional on $(X, W) = (x, w)$, by deconvolution. Hence, the distribution of η given $(X, W) = (x, w)$ is identified. From this distribution, one can identify the function v and the distribution of δ, using the assumption that $v(\widetilde{x}_1, \widetilde{w}_1, \delta) = \delta$, as in Matzkin (2003). Specifically, for any t

$$F_\delta(t) = F_{\eta|(X_1,W_1)=(\widetilde{x}_1,\widetilde{w}_1)}(t)$$

and for any t, x_1, w_1

$$v(x_1, w_1, t) = F^{-1}_{\eta|(X_1,W_1)=(x_1,w_1)}\left(F_{\eta|(X_1,W_1)=(\widetilde{x}_1,\widetilde{w}_1)}(t)\right)$$

Proof of Theorem 4: Following a reasoning similar to that used in the proof of Theorem 2, one can show that

$$\begin{aligned}
& F_{Y^*}(y^*) \\
= \; & F_{Y^*+\eta}(y^*|W_1 = \overline{w}_1) \\
= \; & F_{Y^*+\eta}\left(y^*|W_1 = \overline{w}_1, (X_1, X_2, W_2) = \left(\overline{t} - y^*, \overline{x}_2, \overline{w}_2\right)\right) \\
= \; & \Pr\left(Y^* + \eta \leq y^*|W_1 = \overline{w}_1, (X_1, X_2, W_2) = \left(\overline{t} - y^*, \overline{x}_2, \overline{w}_2\right)\right) \\
= \; & \Pr\left(m\left(Y^* + \eta + X_1, X_2, W_2\right) \leq m\left(\overline{t}, \overline{x}_2, \overline{w}_2\right)|W_1 = \overline{w}_1, (X_1, X_2, W_2) = \left(\overline{t} - y^*, \overline{x}_2, \overline{w}_2\right)\right) \\
= \; & F_{Y|X=\left(\overline{t}-y^*,\overline{x}_2\right),W=(\overline{w}_1,\overline{w}_2)}\left(m\left(\overline{t} - y^*, \overline{x}_2, \overline{w}_2\right)\right) \\
= \; & F_{Y|X=\left(\overline{t}-y^*,\overline{x}_2\right),W=(\overline{w}_1,\overline{w}_2)}(\alpha)
\end{aligned}$$

where the last equality follows by the restriction that $m\left(\overline{t}, \overline{x}_2, \overline{w}_2\right) = \alpha$. Hence, the distribution of Y^* is identified. The function m is identified because then, for any (t, x_1, x_2, w_2),

$$\begin{aligned} m(t, x_2, w_2) &= F^{-1}_{Y|X=(x_1,x_2),W=(\overline{w}_1,w_2)}\left(F_{Y^*}(t-x_1)\right) \\ &= F^{-1}_{Y|X=(x_1,x_2),W=(\overline{w}_1,w_2)}\left(F_{Y|X=\left(\overline{t}-t+x_1,\overline{x}_2\right),W=(\overline{w}_1,\overline{w}_2)}(\alpha)\right)\right) \end{aligned}$$

The rest follows very closely the proof of Theorem 2, and is therefore omitted.

Proof of Theorem 5: Following arguments as in the proof of Theorem 2, we have that since (ξ, δ) is distributed independently of (X, W, Z), ξ is distributed independently of (X_2, W_2, Z) conditional on any values of (X_1, W_1), $\xi + \delta$ is distributed independently of (X_2, W_2, Z) conditional on any values of (X_1, W_1), and $Y^* + \eta = s(Z, \xi) + v(X_1, W_1, \delta)$ is distributed independently of (X_2, W_2) conditional on any values of (X_1, W_1, Z). Moreover, when $(X_1, W_1, Z) = (\overline{x}_1, \overline{w}_1, \overline{z})$, $Y^* + \eta = \xi$; and when $(X_1, W_1) = (\widetilde{x}_1, \widetilde{w}_1)$, $\eta = v(\widetilde{x}_1, \widetilde{w}_1, \delta) = \delta$ and $Y^* + \eta = Y^* + \delta$. Hence, for any e and (x_2, w_2)

$$\begin{aligned} &(T5.1) \quad F_{\xi}(e) \\ &= F_{\xi|X=(x_1,x_2),W=(w_1,w_2),Z=z}(e) \\ &= F_{Y^*+\eta|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=\overline{z}}(e) \\ &= \Pr\left(Y^* + \eta \leq e \middle| (X_1, W_1, Z) = (\overline{x}_1, \overline{w}_1, \overline{z}), (X_2, W_2) = (x_2, w_2)\right) \\ &= \Pr\left(m(Y^* + \eta, X_2, W_2) \leq m(e, x_2, w_2) \,|\, (X_1, W_1, Z) = (\overline{x}_1, \overline{w}_1, \overline{z}), (X_2, W_2) = (x_2, w_2)\right) \\ &= F_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=\overline{z}}\left(m(e, x_2, w_2)\right) \end{aligned}$$

Since for all t, $m(t, \overline{x}_2, \overline{w}_2) = t$

$$(T5.2) \qquad F_\xi(e) = F_{Y|X=(\overline{x}_1,\overline{x}_2),W=(\overline{w}_1,\overline{w}_2),Z=\overline{z}}(e)$$

Hence, F_ξ is identified. Using this in (T5.1), we get that

$$F_{Y|X=(\overline{x}_1,\overline{x}_2),W=(\overline{w}_1,\overline{w}_2),Z=\overline{z}}(e) = F_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=\overline{z}}(m(e, x_2, w_2))$$

Since our assumption imply that $F_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),,Z=\overline{z}}$ is invertible, this implies that for any t and any (x_2, w_2)

$$(T5.3) \qquad m(t, x_2, w_2) = F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=\overline{z}}\left(F_{Y|X=(\overline{x}_1,\overline{x}_2),W=(\overline{w}_1,\overline{w}_2),Z=\overline{z}}(t)\right)$$

Hence, m is identified nonparametrically.

Using a similar reasoning as in (T5.1), we get that for any (x_2, w_2) and any z

$$\begin{aligned} (T5.4) \qquad & F_\xi(e) \\ = \quad & F_{\xi|X=(x_1,x_2),W=(w_1,w_2),Z=z}(e) \\ = \quad & F_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=z}(s(z,e)) \end{aligned}$$

Equation (T5.2) together with the strict monotonicity of $F_{Y^*|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=z}$ imply then that

$$(T5.5) \qquad s(z, e) = F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=z}\left(F_{Y|X=(\overline{x}_1,\overline{x}_2),W=(\overline{w}_1,\overline{w}_2),Z=\overline{z}}(e)\right)$$

Hence, the function s is identified. Since the distribution of ξ is also identified, this implies that the distribution of Y^* conditional on Z is also identified.

The distribution of δ is identified because for any t

$$\begin{aligned} & F_{\xi+\delta}(t) \\ = \; & F_{Y|(X_1,W_1)=(\widetilde{x}_1,\widetilde{w}_1),(X_2,W_2)=(x_2,w_2),Z=\overline{z}}(m\left(t,x_2,w_2\right)) \end{aligned}$$

Hence, using (T5.3), it follows that for any x_2, w_2

$$F_{\xi+\delta}(t) = F_{Y|(X_1,W_1)=(\widetilde{x}_1,\widetilde{w}_1),(X_2,W_2)=(x_2,w_2),Z=\overline{z}}(F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=\overline{z}}\left(F_{Y|X=(\overline{x}_1,\overline{x}_2),W=(\overline{w}_1,\overline{w}_2),Z=\overline{z}}\left(t\right)\right))$$

This implies that the distribution of $\xi+\delta$ is identified. Hence, by deconvolution, we can obtain the distribution of δ, using the already identified distribution of ξ.

The distribution of η conditional on (X_1, W_1) is identified because for any t, x_2, w_2

$$\begin{aligned} & F_{\xi+\eta|(X_1,W_1)=(x_1,w_1)}(t) \\ = \; & (X_1,W_1)=(x_1,w_1)\,,(X_2,W_2)=(x_2,w_2)\,,Z=\overline{z} \\ = \; & F_{Y|(X_1,W_1)=(x_1,w_1),(X_2,W_2)=(x_2,w_2),Z=\overline{z}}(m\left(t,x_2,w_2\right)) \\ = \; & F_{Y|(X_1,W_1)=(x_1,w_1),(X_2,W_2)=(x_2,w_2),Z=\overline{z}}(F^{-1}_{Y|X=(\overline{x}_1,x_2),W=(\overline{w}_1,w_2),Z=\overline{z}}\left(F_{Y|X=(\overline{x}_1,\overline{x}_2),W=(\overline{w}_1,\overline{w}_2),Z=\overline{z}}\left(t\right)\right)) \end{aligned}$$

Hence, since the distribution of ξ is known, and ξ is distributed independently of (X_1, W_1), the distribution of η conditional on (X_1, W_1) is identified. From this conditional distribution and the distribution of δ, we can identify $v\left(x_1, w_1, \delta\right)$, by the arguments in Theorem 1, as

$$v\left(x_1, w_1, \delta\right) = F^{-1}_{\eta|(X_1,W_1)=(x_1,w_1)}\left(F_\delta(\delta)\right)$$

This completes the proof.

Proof of Theorem 6: Assumption A implies that Assumption 2.1 in Matzkin (2005a) is satisfied. Consider the matrix

$$A = \begin{pmatrix} \frac{\partial r_1(y,x)}{\partial y} & \frac{\partial r_1(y,x)}{\partial x} & 0 \\ \frac{\partial \widetilde{r}_1(y,x)}{\partial y} & \frac{\partial \widetilde{r}_1(y,x)}{\partial x} & 0 \\ \frac{\partial \widetilde{r}_2(y,x,z)}{\partial y} & \frac{\partial \widetilde{r}_2(y,x,z)}{\partial x} & \frac{\partial \widetilde{r}_2(y,x,z)}{\partial z} \end{pmatrix}$$

The restrictions that $\partial \widetilde{r}_2(y,x,z)/\partial z \neq 0$ and that the ratio of the derivatives of r_1 is different than the ratio of the derivatives of r_2 imply that the determinant of this matrix is different from 0 at (y,x,z). The assumption on the determinant implies that

$$\begin{aligned} \frac{\partial \log\left|\frac{\partial r(y,x,z)}{\partial y}\right|}{\partial y} - \frac{\partial \log\left|\frac{\partial \widetilde{r}(y,x,z)}{\partial y}\right|}{\partial y} &= \frac{\partial \log\left|\frac{\partial r(y,x,z)}{\partial y}\right|}{\partial x} - \frac{\partial \log\left|\frac{\partial \widetilde{r}(y,x,z)}{\partial y}\right|}{\partial x} \\ &= \frac{\partial \log\left|\frac{\partial r(y,x,z)}{\partial y}\right|}{\partial z} - \frac{\partial \log\left|\frac{\partial \widetilde{r}(y,x,z)}{\partial y}\right|}{\partial z} = 0 \end{aligned}$$

These, together with the assumption on the derivative of $f_{Y^*,\omega}$ imply that Condition 4.1 in Matzkin (2005a) is satisfied. Lemma 4.2 in Matzkin (2005b) imply then that $(\widetilde{r}_1, \widetilde{r}_2)$ is not observationally equivalent to $\left(r_1, r_2\right)$. Hence, by the definition of observational equivalence, the result follows.

7. References

BATTISTIN, E. (2003) "Errors in Survey Reports of Consumption Expenditures," working paper # 0307, *Institute for Fiscal Studies*, London.

BOLLINGER, C.R. (1998) "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics,* Vol. 16, No. 3, pp. 576-594.

BOUND, J., C. BROWN, and N. MATHIOWETZ (2001) "Measurement Error in Survey Data," in *Handbook of Econometrics,* Vol. 5, edited by J.J. Heckman and E. Leamer, 3705-3843. Amsterdam: Elsevier.

BOUND, J. and A. KRUEGER (1991) "The Extent of Measurement Error in Longitudinal Earning Data: Do Two Wrongs Make a Right?," *Journal of Labor Economics*, 16, 576-94.

BOUND, J., C. BROWN, and N. MATHIOWETZ (2001) "Measurement Error in Survey Data," in *Handbook of Econometrics,* Vol. 5, edited by J.J. Heckman and E. Leamer, 3705-3843. Amsterdam: Elsevier.

BRIESCH, R., P. CHINTAGUNTA, and R.L. MATZKIN (2005) "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," mimeo, Northwestern University.

CARROLL, R.J., D. RUPPERT, and D. STEFANSKI (1995) *Measurement Error in Nonlinear Models,* New York: Chapman and Hall.

CHEN, X, H. HONG and E. TAMER (2005) “Measurement Error Models with Auxiliary Data,” *Review of Economic Studies,* 72,2, 343-366.

CHESHER, A. (1991) "The Effect of Measurement Error," *Biometrika,* 78, 451.

CHESHER, A. (1998) "Polynomial Regression with Covariate Measurement Error," Discussion Paper 98/448, University of Bristol.

CHESHER, A. (2003) "Identification in Nonseparable Models," *Econometrica,* 71, 1401-

1444.

HAUSMAN, J., W. NEWEY, AND J. POWELL (1995) “Nonlinear Errors in Variables. Estimation of Some Engel Curves,” *Journal of Econometrics,* 65, 205-233.

HAUSMAN, J., W. NEWEY, H. ICHIMURA, and J. POWELL (1991) “Measurement Errors in Polynomial Regression Models,” *Journal of Econometrics,* 50, 273-295.

HOROWITZ, J.L. and C.F. MANSKI (1995) "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica,* 63, 2, 281-302.

HURD, M.D., D. McFADDEN, H. CHAND, L. GAN, A. RMERRILLL and M. ROBERTS (1998) "Consumption and Saving Balances of the Elderly: Experimental Evidence on Survey Response Bias", in *Frontiers in the Economics of Aging,* edited by D. Wise, 353-387, Chicago, IL: University of Chicago Press.

HSIAO, C. (1989) “Consistent Estimation for Some Nonlinear Errors-in-Variables Models,” *Journal of Econometrics,* 41, 159-185.

HSIAO, C. and Q. WANG (2000) “Estimation of Structural Nonlinear Errors-in-Variables Models by Simulated Least Squares Method,” *International Economic Review,* 41, 523-542.

IMBENS, G.W. AND W.K. NEWEY (2003) “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” mimeo, UCLA.

LEWBEL, A. (1998) “Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica,* 66, 105-121.

LI, T. (2002) “Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models,” *Journal of Econometrics,* 110, 1-26.

LI, T. and Q. VUONG (1998) "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis,* 65, 139-165.

McFADDEN, D. (1994) "Contingent Valuation and Social Choice," *American Journal of Agricultural Economics,* 76, 689-708.

McFADDEN, D., N. SCHWARZ, and J. WINTER (2003) "Measuring Perceptions and behavior in Household Surveys," mimeo, Mannheim Research Institute for the Economics of Aging.

MANSKI, C.F. (2005) "Partial Identification in Econometrics," forthcoming in *The New Palgrave Dictionary of Economics,* 2nd Edition, London: McMillan.

MATZKIN, R.L. (1993) "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometics,* 58.

MATZKIN, R.L. (1994) "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics,* Vol. 4, edited by R.F. Engel and D. McFadden.

MATZKIN, R.L. (2003) "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.

MATZKIN, R.L. (2004) "Unobservable Instruments," mimeo, Northwestern University.

MATZKIN, R.L. (2005a) "Identification in Nonparametric Simultaneous Equations," mimeo, Northwestern University.

MATZKIN, R.L. (2005b) "Estimation in Nonparametric Simultaneous Equations," mimeo, Northwestern University.

MOLINARI, F. (2005) "Partial Identification of Probability Distributions with Misclassified Data," mimeo, Cornell University.

NEWEY, W. (2001) Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *Review of Economics and Statistics*, 83, 616-627.

PHILIPSON, T. (1997) "Data Markets and the Production of Surveys," *Review of Economic*

Studies, 64 (1), 47-72.

PHILIPSON, T. (2001) "Data Markets, Missing Data, and Incentive Pay," *Econometrica,* 69 (4), 1099-1111.

POTERBA, J. and L. SUMMERS (1986) "Reporting Errors and Labor Market Dynamics," *Econometrica,* 54 (6), 1319-1338.

SCHENNACH, S.M. (2004) "Estimation of Nonlinear Models with Measurement Error," *Econometrica,* 72, 1, 33-75.

SCHENNACH, S.M. (2005) "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," mimeo, University of Chicago.

SCHWARZ, N., H.J. HIPPPLER, B. DEUTSCH, and F. STRACK (1985) "Response Categories: Effects on Behavioral Reports and Comparative Judgements," *Public Opinion Quarterly*, 49, 388-395.

TOURANGEAU, R., L.J. RIPS, and K. RASINSKI (2000) *The Psychology of Survey Response.* New York, NY and Cambridge, UK: Cambridge University Press.

WANG, L. and C. HSIAO (1995) "Simulation-Based Semiparametric Estimatin of Nonlinear Error-in-Variables Models," Working Paper, University of Southern California.