Journal of Economic Literature 2011, 49:4, 901–937
http:www.aeaweb.org/articles.php?doi=10.1257/jel.49.4.901

Nonlinear Models of Measurement Errors

Xiaohong Chen, Han Hong, and Denis Nekipelov*

Measurement errors in economic data are pervasive and nontrivial in size. The presence of measurement errors causes biased and inconsistent parameter estimates and leads to erroneous conclusions to various degrees in economic analysis. While linear errors-in-variables models are usually handled with well-known instrumental variable methods, this article provides an overview of recent research papers that derive estimation methods that provide consistent estimates for nonlinear models with measurement errors. We review models with both classical and nonclassical measurement errors, and with misclassification of discrete variables. For each of the methods surveyed, we describe the key ideas for identification and estimation, and discuss its application whenever it is currently available. (*JEL* C20, C26, C50)

1. *The Importance of Measurement Errors*

The primary purpose of this review is to introduce readers to recent work on measurement error in the nonlinear context. We aim to achieve this purpose by offering intuitions about the consequences of possible measurement errors in key variables and by understanding the circumstances under which consistent estimation of the parameters of interests is possible.

Measurement errors in economic data are pervasive and nontrivial in size. The presence of measurement errors causes biased and inconsistent parameter estimates and leads to erroneous conclusions to various degrees in economic analysis. The importance of measurement errors in analyzing the empirical implications of economic theories is highlighted in Milton Friedman's seminal book on the consumption theory of the permanent income hypothesis (Friedman 1957). In Friedman's model, both consumption and income consist of a permanent component and a transitory component that can arise from measurement errors or genuine fluctuations. The marginal propensity to consume relates the permanent component of consumption to the permanent income component. Friedman shows that, because of the attenuation bias, the slope coefficient of a regression of observed consumption on observed income would lead to an underestimate of the marginal propensity to consume.

* Chen: Yale University. Hong: Stanford University. Nekipelov: University of California, Berkeley. The authors acknowledge generous research support from the National Science Foundation and Tim Armstrong for research assistance. The authors thank Roger Gordon and the five anonymous referees for comments that greatly improve the quality of the survey. The usual disclaimer applies.

The empirical importance of measurement errors is also investigated in Andrew Chesher and Christian Schluter (2002), who studied how measurement errors of various degree can lead to bias in the estimation of poverty indexes and Gini coefficients. Their application to regional poverty and inequality comparisons in India shows that, while measurement errors of plausible magnitudes can seriously disturb the view of regional differences in inequality when comparisons are made within rural or urban areas, the differences in measured poverty and inequality across rural and urban areas cannot plausibly be explained by differences in measurement error.

When measurement errors are independent of the underlying true variables, they are called *classical*. Classical measurement errors can introduce substantial biases. However, the commonly asserted perception that classical measurement errors result in an attenuation bias is only valid under the strong assumptions of an univariate linear regression model. The presumption of attenuation bias no longer holds in multivariate linear regression models, even if only one of the multiple regressors is subject to measurement errors. For linear models subject to measurement errors, consistent estimates are possible with a second independent noisy measure for each variable with measurement error, or with a proxy variable that represents the expectation of the true variable for a given information set.

Further problems arise when errors are not independent of the true values, as found in various studies making use of auxiliary data with observations free of the measurement error. With such auxiliary data, though, consistent estimates are still feasible.

The implication of linear models does not carry over directly when the model is nonlinear. The nature of the bias depends on the curvature of the nonlinear function, and there is no longer a clear presumption of attenuation bias even in univariate models except perhaps when errors are small. While the use of instrumental variables (IV) methods, proxy variables, or auxiliary samples are still a common starting point even with nonlinear models, the specific conditions for identification and estimation do not carry over directly. What does carry over from the linear setting to nonlinear models is the need of using extra information in the form of either instrumental variables or auxiliary samples in order to overcome the loss of information induced by the presence of measurement errors. The intuition from linear models can potentially lead us astray in nonlinear models. See for example, Raymond J. Carroll and Leonard A. Stefanski (1990), William T. Dickens and Brian A. Ross (1984), and Gene T. Hwang and Stefanski (1994).

While linear errors-in-variables (EIV) models are usually handled with well-known instrumental variable methods, in the rest of this article we provide an overview of recent research papers that derive estimation methods that provide consistent estimates for nonlinear models with measurement errors. We mainly describe key ideas for identification and estimation, and discuss applications of these new methods whenever they are currently available.

The rest of the survey is organized as follows. Section 2 briefly summarizes the different approaches to various types of measurement error models. Section 3 reviews results on nonlinear EIV models with classical measurement errors. Section 4 presents recent results on nonlinear EIV models with nonclassical measurement errors, including misclassification in models with discrete variables. Section 5 on contaminated and corrupted data surveys papers that develop set identification methods under a different mechanism that generates measurement errors in the data. Section 6 briefly concludes.

2. *Varieties of Measurement Error Models*

The goal of most measurement error models is about the estimation of the conditional mean regression function:

$$E(y^* | x^*) \;=\; g(x^*) \tag{1}$$

where some elements of (x^*, y^*) might not be observed. The conditional mean function $g(x^*)$ can be specified either parametrically, nonparametrically, or semiparametrically. While the rest of this survey will focus mainly on nonlinear parametric models, in some settings it can be easier to understand identification arguments for nonparametric models. Parametric examples of $g(x^*)$ include the exponential regression $g(x^*) = \exp(x^*\beta)$ often used in count data analysis, and the probit binary response model in which $g(x^*) = \Phi(x^*\beta)$. While most of our survey applies to the conditional model, in a few cases we will also study the unconditional moment that takes the form of

$$Em(y^*, \beta) \;=\; 0. \tag{2}$$

While y^* denotes the *true* unobserved variables of the model, y is the observed variable that provides the noisy measurement for y^*.

A general assumption that will be maintained throughout the survey for the conditional model (1) is that conditional on the true unobserved variable x^*, the measurement error contained in x does not provide additional information regarding the conditional distribution of the dependent outcome variable. This assumption also holds when multiple mismeasured proxies of x^* are available, for example if w is a second measurement of the true x^*. Formally,

Assumption 2.1
$F(y^* | x^*, x, w) \;=\; F(y^* | x^*, x) \;=\; F(y^* | x^*)$.

In the rest of this section and in section 3 where the classical errors-in-variables assumption is maintained, it is also assumed that the errors in measuring y^* are purely additive independent noise: $E(y | y^*, x^*, x) = y^*$. Together with assumption 2.1, this implies that

$$E(y | x^*, x) \;=\; E(y^* | x^*, x) \;=\; E(y^* | x^*). \tag{3}$$

The information about model parameters has to be inferred from the joint distribution of the observed variables x and y, which given (3) satisfies the following relation:

$$\begin{aligned} E(y | x) \;&=\; \int E(y^* | x, x^*) f(x^* | x) dx^* \\ &=\; \int g(x^*, \beta) f(x^* | x) dx^*. \end{aligned} \tag{4}$$

With no loss of generality, let $x = x^* + \eta$. Identification of the model parameters β hence depends crucially on the functional form of $g(x^*, \beta)$ and on the available knowledge of the relation between x^* and x, or equivalently the relation between η and x^*. Measurement error models can be differentiated according to the assumptions they impose on the distribution of η given x^* and the method they use to recover information about the distribution of η.

While the remaining sections of this survey will focus on nonlinear models, a quick review of the linear least square regression model $y_i^* = x_i^{*\prime}\beta + u_i$ is in order here. The classical measurement error assumption maintains that the measurement errors in any of the variables in the data set are independent of all the true variables. Under this assumption, measurement errors in the dependent variable $y_i = y_i^* + \varepsilon_i$ do not lead to inconsistent estimate of the regression coefficients. The main consequence of the presence of measurement errors in the dependent variables is that they inflate the standard errors of these regression coefficient estimates. If the measurement errors have nonzero means, they will also shift the intercept even when the true slope is zero. On the other hand, Ragnar Frisch (1934) demonstrated that independent errors that are present in the observations of the regressors $x_i = x_i^* + \eta_i$ lead to attenuation bias in a simple univariate regression model and to inconsistent regression coefficient estimates in general. In nonlinear models, Zvi Griliches and Vidar Ringstad (1970) demonstrate that the bias introduced by measurement errors can be more substantial than the bias in linear models. For both univariate and multivariate nonlinear models, it is in general not possible to sign the direction of bias introduced by the presence of measurement errors, which typically depends on the curvature of the nonlinear regression function.

Measurement error models can be regarded as a special case of models with endogenous regressors; the linear regression model can be consistently estimated using the method of IV. A valid instrumental variable often comes from a second measurement of the error-prone true variable: $w_i = x_i^* + v_i$, which is subject to another independent measurement error v_i. Because of linearity, it is sufficient for v_i and η_i to be uncorrelated with x_i^* and u_i and for v_i to be uncorrelated with η_i, which is a weaker condition than requiring full independence.

The use of a second measurement as an instrumental variable indicates the extra information one needs to identify the model parameters. In the simple univariate regression model above, under the additional assumption that $E(x^*|x)$ is a linear function,

$E(y^*|x) = \beta\frac{\sigma_{x^*}^2}{\sigma_x^2}x$, where $E(x^*|x) = \frac{\sigma_{x^*}^2}{\sigma_x^2}\,x$ is

obviously a feature of the distributions of x^* and η. The second measurement w is used to identify $\sigma_{x^*}^2$ and therefore the features of the distributions of x^* and η that are needed to identify β. In the linear model, only the second moments of x and w are needed for identification. The instrumental variable method does not even require the linear conditional mean assumption. However, in more complex nonlinear models, nonparametric methods are usually needed to estimate the entire joint distribution of x^* and η without making restrictive assumptions.

Even in linear models, identification is difficult without additional information provided by a second measurement or an instrumental variable. Olav Reiersøl's (1950) identification analysis of the bivariate linear EIV model without additional instruments shows that identification results are limited. Using a characteristic function approach, Reiersøl (1950) shows that the slope coefficient is identified if and only if the observed variables are not normally distributed. The intercept parameter and the variance matrix of the measurement errors are identified only if the components of the true variables are not divisible by a normal distribution, and one of the measurement errors is identically equal to zero. He also shows that if the measurement errors are independent, the slope coefficient is identified given certain conditions on the curvature of the characteristic function of the unobserved true variables. However, Arthur Lewbel (1997) shows that a second measurement, or an instrumental variable, can often be difficult to find in many empirical applications.

Even when the coefficient in the linear EIV model is not identified by additional information or stronger assumptions, it can still be bounded between a forward regression estimate and an inverse regression estimate (Corrado Gini 1921 and Frisch 1934). This so-called Frisch bound is also generalized to identified sets in multivariate regressions in Steven Klepper and Edward E. Leamer (1984).

An alternative to the classical measurement error assumption is the *proxy assumption* which states that η is independent of x instead of x^* in $x = x^* + \eta$. Under the *proxy* assumption, the usual univariate linear regression of y on x will consistently estimate the coefficient β. It leads to consistent estimates in multivariate models only if η is also uncorrelated with the other perfectly observed regressors. Obviously, the proxy assumption is not compatible with the classical measurement error assumption. Which assumption is valid is ultimately an empirical issue.

Bias in nonlinear models is more complex than the attenuation bias in linear models. Using a small error expansion, Chesher (1991) obtained an approximation formula to characterize the bias in a general nonlinear regression model when the regressor is measured with error and the measurement error distribution is not necessarily normal. The approximate bias depends on the derivatives of the regression function with respect to the mismeasured regressor and the curvature of the distribution functions of the true regressor and the mismeasured regressors.

Chesher, Montezuma Dumangane, and Richard J. Smith (2002) make use of the small measurement error approximation method to develop a statistical test for the presence of measurement errors in a nonlinear duration response model, and find that the presence of measurement errors can severely bias the direction of duration dependence. Duration dependence can be biased by the presence of unobserved heterogeneity, as well as by the presence of nonlinear measurement errors.

IV methods for linear errors-in-variables models are usually not applicable for nonlinear models. Yasuo Amemiya (1985) finds that instrumental variable estimators are only consistent for nonlinear errors-in-variables models when the order of magnitude of measurement errors decreases to zero as the sample size increases to infinity, and gives conditions of the rate of decrease of the measurement errors. In general, without the "small measurement error" assumption of Amemiya (1985), instrument variable methods are inconsistent for nonlinear EIV models. Intuitively, linear instrumental variable methods make use of second moments of the observed regressors and instruments to identify parameters in a linear regression relation. In nonlinear models, identification typically depends on the entire joint distribution of the observed data, which is not recovered in conventional linear instrumental variable methods.

Analogous to linear models, the choice between different identification and estimation strategies for nonlinear models depends on the information that is available in the data. Under the classical measurement error assumption, if a second independent noisy observation is available, deconvolution techniques can be used to estimate the measurement error distribution by picking out the shared true variation in the underlying series. The implementation details of the deconvolution method depend on the structure of the regression function and the measurement error distribution, and on whether the nonlinear model is parametric or nonparametric. These are described in detail in subsections 3.1 to 3.3. The strong instrumental variable method used in subsection 3.4, on the other hand, bears some resemblence to the proxy assumption used in linear models.

The assumption of classical measurement errors can fail in both linear and nonlinear models for a variety of reasons. Recent empirical evidence demonstrates that, in some survey data, the amount of misreporting can be correlated with the level of the true variables. Without explicit knowledge of the correlation structure, a validated subsample is typically required to document the nature of the measurement error and to allow for appropriate corrections in the full sample. For example, John Bound and Alan B. Krueger (1991) compare reported income in the 1977 wave of the Current Population Survey to the social security match file for respondents who provide a valid social security number. They find that measurement error is correlated with the true variable, especially in the tails of the distribution. Bound et al. (1994) and Christopher R. Bollinger (1998) use a validation study on one large firm from the Panel Study of Income Dynamics and find evidence against the classical measurement error model and especially against the independence assumption between the measurement error and the true variable.

The classical measurement error assumption will also necessarily fail in discrete data, for which the problem of mismeasurement takes the form of misclassification. In discrete data, model identification often comes from the exogenous variation of an instrumental variable such as a second measurement. In the context of a linear wage regression model with categorical regressors of educational achievement, Thomas J. Kane, Cecilia Elena Rouse, and Douglas Staiger (1999) relax the classical EIV assumption by specifying a linear functional form for the conditional expectations of the two measurements on the unobserved true variables. Nonzero intercept coefficients and differences of the slope coefficients from one are interpreted as evidences of nonclassical measurement errors. Kane, Rouse, and Staiger (1999) show that the parameters in the linear regression that relates the observed measurement to the unobserved true variable can be jointly identified with the coefficient parameters in the wage regression model from the joint second moments of the two observed reports of education attainments and the observed log wages. Their method depends crucially on the linearity assumptions both in the wage equation and in the measurement equation. Section 4 surveys recent methods that relax these assumptions and are suitable for nonlinear models.

If strong identifying information such as an instrumental variable or an auxiliary sample is not available, only bounds can be placed on the identified set of parameters. Section 5 reviews these techniques for contaminated and corrupted data. This is a particular measurement error mechanism that assumes that some data are correctly observed while little information is available on how the other misreported data is related to the underlying truth.

While we shall focus on cross-sectional data and maintain the assumption that the data are independently and identically distributed, measurement error problems can be equally severe in time series and panel data sets. Reviews of earlier results on this subject can also be found in Wayne A. Fuller (1987), Carroll et al. (2006), Tom Wansbeek and Erik Meijer (2000), Bound, Charles Brown, and Nancy Mathiowetz (2001), Jerry A. Hausman (2001), and Geert Ridder and Robert Moffit (2007), among others.

3. *Nonlinear EIV Model with Classical Errors*

Identification in general nonlinear EIV models requires stronger assumptions than in linear EIV models. Amemiya (1985) shows that the standard IV assumption (i.e., the instrument is correlated with the mismeasured variable but not correlated with the measurement error) is no longer sufficient for identification of nonlinear EIV regression models under the classical additive independent error assumption.

All the methods for nonlinear EIV models with classical measurement errors presented in this section are applications of the *deconvolution* method. The convolution of two functions f and g is a function $f^* g$ such that $(f^* g)(x) = \int f(u) g(x - u)\, du$. Under the classical EIV assumption, the distribution of y_i is a convolution of the distributions of y_i^* and ε_i: $f(y_i) = \int f^*(y_i - \varepsilon_i) f_\varepsilon(\varepsilon_i)\, d\varepsilon_i$. The deconvolution method recovers the distribution of $y_i^* \in \mathbb{R}^k$ from the observed distribution of y_i and information about the distribution of ε_i in $y_i = y_i^* + \varepsilon_i$. It makes extensive use of the characteristic function of the error ε, which is denoted $\phi_\varepsilon(t)$ and is defined as $\phi_\varepsilon(t) = Ee^{it'\varepsilon_i} = F[f_\varepsilon](t)$.

Because the density of the observable variable is the convolution of the density of the unobservable variable and the density of the measurement error, the characteristic function of the observable variable is the product of characteristic function of the unobservable variable and the measurement error: $\phi_y(t) = \phi_{y^*}(t) \times \phi_\varepsilon(t)$. Therefore, if the distribution function of ε and hence its characteristic function is known or can be recovered from information available in the sample, then the characteristic function of y_i^* can be recovered from the ratio of the characteristic functions $\phi_y(t)$ and $\phi_\varepsilon(t)$ of y_i and ε_i: $\phi_{y^*}(t) = \phi_y(t)/\phi_\varepsilon(t)$. Once the characteristic function of y^* is identified, its density can be recovered from the inverse Fourier transformation of the corresponding characteristic function

$$f(y^*) = \left(\frac{1}{2\pi}\right)^k \int \phi_{\mathbf{y}^*}(\mathbf{t}) e^{-iy^{*\prime}\mathbf{t}} d\mathbf{t}.$$

Subsequently, the latent moment condition can be identified for each possible parameter value as $Em(y^*; \beta) = \int m(y^*; \beta) f(y^*) dy^*$. Consistent estimators can then be formulated as the sample analogs of the population identification procedure. This approach is reviewed in section 3.2.

Deconvolution techniques are widely used in the statistics literature which typically assumes that the measurement error distributions are completely known. See Carroll and Peter Hall (1988), Jianqing Fan (1991), and Fan and Young K. Truong (1993) for the optimal convergence rates for nonparametric deconvolution problems. Marie-Luce Taupin (2001) and Cristina Butucea and Taupin (2008) provide results for semiparametric estimation.

3.1 *Nonlinear EIV Models with Polynomial Structure*

3.1.1 *Polynomial Inverse Error Characteristic Function*

When the inverse of the characteristic function of the measurement error $\phi_\varepsilon(\mathbf{t})$ is a parameterized polynomial, its parameters can be estimated jointly with the parameter of the econometric model β. For example, the characteristic function of independent Laplace (double exponential) errors $\varepsilon \in \mathbb{R}^k$ takes the form of $\phi_\varepsilon(\mathbf{t}) = \Pi_{j=1}^k \left(1 + ½\sigma_j^2 t_j^2\right)^{-1}$. When the measurement errors are Laplace (double exponential) with zero means and unknown variances and the measurement errors are independent of the latent variables and are independent of each other. Han Hong and Elie Tamer (2003) show that the moment condition for the latent random vector $\mathbf{y}^*$, $Em(\mathbf{y}^*; \beta)$, can be translated into the moment condition for the observable random variable $\mathbf{y}$ as $Em(\mathbf{y}^*; \beta) = Em(\mathbf{y}; \beta, \sigma)$, for

$$m(\mathbf{y}; \beta, \sigma) = m(\mathbf{y}; \beta) + \sum_{l=1}^{k} \left(-\frac{1}{2}\right)^l \sum_{1\le j_1 < \cdots < j_l \le k} \cdots \sum \sigma_{j_1}^2 \cdots \sigma_{j_l}^2 \times \frac{\partial^{2l}}{(\partial y_{j_l})^2 \cdots (\partial y_{j_l})^2} m(\mathbf{y}; \beta).$$

In the above, we have used the notation $\frac{\partial^{2l}}{(\partial y_{j_1})^2 \cdots (\partial y_{j_l})^2}$ to denote the $2l$th partial derivative twice with respect to each of y_{j_1} to y_{j_l}.

The sample analog of the revised moment condition can be used to obtain point estimates of both the parameter of the econometric model β and the parameters characterizing the distribution of the measurement error $\sigma \equiv \{\sigma_j, j = 1, \ldots, k\}$.

Consider for example the following model: $E[y|x^*] = g(x^*; \beta)$, $x = x^* + \eta$, where $g(\cdot;\cdot)$ is a known twice differentiable function and x^* is a latent variable defined on $\mathbb{R}$ such that the conditional variance $\mathrm{var}(y|x^*)$ is finite. This model implies the unconditional moment restriction, $E[\mathbf{h}(x^*)(y - g(x^*; \beta))] = 0$ for an $h \times 1$ $(h > \dim(\beta))$ vector of measurable functions $\mathbf{h}(\cdot)$. Then, the revised moment conditions in terms of observed variables are

$$
\begin{aligned}
E\Big[\mathbf{h}(x)(y - g(x;\beta))& \\
- \frac{1}{2}\sigma^2\Big(\mathbf{h}^{(2)}(x)y \; - \; & \mathbf{h}^{(2)}(x)g(x;\beta) \\
&- \; 2\mathbf{h}^{(1)}(x)g^{(1)}(x;\beta) \\
&- \; \mathbf{h}(x)g^{(2)}(x;\beta)\Big)\Big] \; = \; 0.
\end{aligned}
$$

In the above, the superscripts in parentheses refer to the derivatives with respect to x. For each candidate parameter value β, the left hand side of the revised moment conditions can be estimated from the sample analog by replacing the expectation with the empirical sum. The usual generalized method of moments (GMM) estimates of β and σ that minimize distance of these moment conditions to zero are consistent and asymptotically normal.

Even if the revised moment condition $E[m(\mathbf{y}; \beta, \sigma)] = 0$ cannot point identify the parameter β, it still contains useful information about β that can be exploited using the information about $\sigma_1^2, \ldots, \sigma_k^2$. The variance of the measurement errors should be smaller than the variance of the "signal" $0 \leq \sigma_j^2 \leq \sigma_{y_j}^2$, where $\sigma_{y_j}^2$ is the variance of the observed random variable y_j. Hong and Tamer (2003) describe how bounds on the variance of the measurement errors can be translated into bounds on the model parameters. In the simplest case of a linear model, the revised moment condition is identical to the original moment condition, and directly identifies the model parameter β. The variance of the measurement error is not identified in the revised moment condition, but can be bounded by the variance of the observed random variable. Whether this model is ultimately relevant for empirical work depends on the strength of evidence supporting the Laplace error assumption. It is an open research question to develop empirical tests for its validity.

3.1.2 *Polynomial Moment Conditions*

Hausman et al. (1991) generalized the instrumental variable method for linear regression models to polynomial regression models in which the regressors are polynomial functions of the error-prone variables. We illustrate their results using a simplified version of the polynomial regression model that they considered: $y = \sum_{j=0}^{K} \beta_j (x^*)^j + r'\phi + u$. The vector r is precisely observed but the scalar x^* is only observed with classical errors. Two measurements of x^*, x and w, are observed which satisfy $x = x^* + \eta$ and $w = x^* + v$. Assume that u, η and v are mutually independent and they are independent of all the true regressors in the model. The independence assumption can be weakened to certain conditional moment independence assumptions. The identification argument is first presented for the simpler case when $\phi = 0$, and then extended to the more general case when $\phi \neq 0$. In both cases, latent cross moments that are needed to compute the normal equation that defines the regression coefficients are computed from the observed cross moments that involve both noisy measurements.

In the first case, when $\phi = 0$, identification of β can be based upon population moments $\xi_j \equiv E(y(x^*)^j)$, $j = 0, \ldots, K$ and $\zeta_m \equiv E(x^*)^m$, $m = 0, \ldots, 2K$, which are the elements of the population normal equations for solving for β. Except for ξ_0 and ζ_0, these moments depend on x^* which is not observed, but they can be solved from the moments of observable variables Exw^j, Ew^j for $j = 0,\ldots,2K$ and $Eyw^j, j = 0,\ldots,K$. Define $\nu_k = Ev^k$. The main idea is that the unobserved moments can be solved from a system of equations defining the observed moments of Exw^j, Ew^j and Eyw^j. In particular, the observable moments satisfy the following relations:

$$(5)\quad Exw^j = E(x^* + \eta)(x^* + v)^j = \sum_{l=0}^{j} \binom{j}{l} \zeta_{l+1} \nu_{j-l}, \quad j = 1, 2K-1,$$

$$(6)\quad Ew^j = E(x^* + v)^j = E \sum_{l=0}^{j} \binom{j}{l} (x^*)^l v^{j-l} = \sum_{l=0}^{j} \binom{j}{l} \zeta_l \nu_{j-l}, \quad j = 1,\ldots,2K,$$

$$(7)\quad Eyw^j = Ey(x^* + v)^j = E \sum_{l=0}^{j} \binom{j}{l} y (x^*)^l v^{j-l} = \sum_{l=0}^{j} \binom{j}{l} \xi_l \nu_{j-l}, \quad j = 1,\ldots,K.$$

Since $\nu_1 = 0$, there are a total of $(5K-1)$ unknowns in $\zeta_1,\ldots,\zeta_{2K}$, $\xi_1,\ldots,\xi_K$ and $\nu_2,\ldots,\nu_{2K}$. Equations (5), (6), and (7) give a total of $5K-1$ equations that can be used to solve for these $5K-1$ unknowns. In particular, the $4K-1$ equations in (5) and (6) jointly solve for $\zeta_1,\ldots,\zeta_{2K}$, $\nu_2,\ldots,\nu_{2K}$. Subsequently, given knowledge of these ζ's and ν's, ξ's can then be recovered from equation (7). Finally, these identified quantities of $\xi_j, j = 0,\ldots,K$ and ζ_m, $m = 0,\ldots,2K$ can be used to recover the parameters β from the normal equations $\xi_l = \sum_{j=0}^{K} \beta_j \zeta_{j+l}$, $l = 0,\ldots,K$.

When $\phi \neq 0$, Hausman et al. (1991) note that the normal equations for the identification of β and ϕ can be based on a second set of moments Eyr, Err' and $Er(x^*)^j, j = 0,\ldots,K$, in addition to the first set of moments ξs and ζs. Since Eyr and Err' can be directly observed from the data, it only remains to identify $Er(x^*)^j, j = 0,\ldots,K$, which can be solved recursively from the following system of equations:

$$Erw^j = Er(x^* + v)^j = E \sum_{l=0}^{j} \binom{j}{l} r (x^*)^l v^{j-l} = \sum_{l=0}^{j} \binom{j}{l} (Er(x^*)^l) \nu_{j-l}, j = 0,\ldots,K.$$

In a follow-up paper, Hausman, Whitney K. Newey, and James L. Powell (1995) apply the identification and estimation methods proposed in Hausman et al. (1991) to estimation of Engel curve specified in the Gorman form using 1982 Consumer Expenditure Survey data set. They found a strong indication of the importance of measurement errors. The estimated variance of measurement error represents 41 percent of the total variance of the logarithm of the measured expenditure. Evidence that higher order terms should be included in the Engel curve specification are found by a test for linearity which has a marginal significance level of 0.04. Strong evidence for measurement errors is also found by a Hausman test that the difference between ordinary least squares and the consistent estimators is statistically significant. The elasticity estimates obtained by the consistent estimation method are also closer to the elasticities computed from the budget share tables.

The polynomial regression model illustrates that only in very strict situations can the moments of the observed variables be used to identify parameters in a nonlinear errors-in-variables model. It is not difficult to conceive of other possible special cases for which the moments of observable variables are sufficient for identification. For example, in a probit model in which both the unobserved true regressors and the measurement errors are jointly normally distributed, the asymptotic bias can be characterized as functions of the second moments of the joint normal distribution, and can be corrected with the knowledge of a second independent measurement. However, in general, without these strict assumptions for special cases, nonlinear errors-in-variables models require estimating the entire joint distribution of the true variables and the measurement errors. The next two subsections provide details. It should also be cautioned that higher moments may not be estimated very accurately in finite samples. Therefore a very large sample is needed in order for the polynomial measurement error regression model to work well.

3.2 *General Nonlinear Models with Double Measurements*

Often the distribution of the measurement errors, or its characteristic function, might not be known and does not have any particular parametric structure. However, if two independent measurements of the latent true variable with additive errors are observed, they can be used to obtain an estimate of the measurement error distribution. Without assuming functional forms of the measurement error distributions, the characteristic function of the true unobserved variable is recovered through information from a second measurement.

Tong Li (2002) provides one method to do this by making use of an identification result due to Kotlarski that is reported in P. Rao (1992). He considers a nonlinear regression model, $y = g(\mathbf{x}^*; \beta) + u$, where $\mathbf{x}^* = (x_1^*, \ldots, x_K^*) \in \mathbb{R}^K$ is the unobservable true regressor and u is the independent random disturbance, with $Eu = 0$, $E(u^2) = \sigma_0^2$. Two measurements $\mathbf{x}$ and $\mathbf{w}$ for $\mathbf{x}^*$ are observed:

$$\mathbf{x} = \mathbf{x}^* + \eta, \quad \mathbf{w} = \mathbf{x}^* + v,$$

$$E(\eta) = E(v) = 0,$$

with individual elements x_k, w_k, η_k, and v_k for $k = 1, \ldots, K$. The measurement errors (η, v) and the unobservable vector of regressors $\mathbf{x}^*$ are mutually independent. Individual components of (η, v) are also independent of each other. In addition, u has zero mean conditional on the latent regressors $\mathbf{x}^*$ and the errors η. Furthermore, Li (2002) assumes that the mismeasured regressors are continuously distributed, and that the characteristic functions of the components of the latent regressor $\mathbf{x}^*$ and the measurement errors η and v are not equal to zero for any finite value of their arguments. This assumption restricts the distributions of measurement errors from decaying "too fast" at infinity.

The conditional mean independence of random disturbance u from the latent regressor $\mathbf{x}^*$ and the errors η implies that the conditional expectation of the dependent variable y given the knowledge of both the latent vector $\mathbf{x}^*$ and the observed $\mathbf{x}$ is determined solely by the function $g(\cdot)$, i.e., $E(y \mid \mathbf{x}^*, \mathbf{x}) = g(\mathbf{x}^*, \beta)$. From this expression, we can obtain the expressions for the conditional expectation of the dependent variable given the observable measurements from the conditional distribution of the latent variable given the observed mismeasured variable (which is determined by the distribution of the classical measurement error). In particular,

$$E(y \mid \mathbf{x}) = E[E(y \mid \mathbf{x}^*, \mathbf{x}) \mid \mathbf{x}] = E[g(\mathbf{x}^*; \beta) \mid \mathbf{x}]$$
$$= \int g(\mathbf{x}^*; \beta) f_{\mathbf{x}^* \mid \mathbf{x}}(\mathbf{x}^* \mid \mathbf{x}) d\mathbf{x}^*.$$

In the above, the second equality follows from the conditional mean independence of u and η given $\mathbf{x}^*$. Therefore if one can obtain a nonparametric estimate $\hat{f}_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*|\mathbf{x})$ of the conditional distribution of the latent variable given the observable mismeasured variable $f_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*|\mathbf{x})$, then one can run a nonlinear regression of y on $\int g(\mathbf{x}^*;\beta)\hat{f}_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*|\mathbf{x})d\mathbf{x}^*$ to obtain a consistent estimate of β. Here is where the classical errors-in-variables assumption comes into play.

To identify $f_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*|\mathbf{x})$, the conditional distribution of the latent variable given the observed measurement, Li (2002) starts by showing that the probability density functions of x_k^* and η_k can be uniquely determined from the joint distribution of (x_k, w_k). To begin with, the joint characteristic function of the mismeasured variables (x_k, w_k) is identified by definition as $\psi_k(t_1, t_2) = Ee^{it_1x_k+it_2w_k}$. Then the characteristic functions for the components of the latent vector and the measurement errors x_k^*, η_k, and v_k, denoted $\phi_{x_k^*}(t)$, $\phi_{\eta_k}(t)$ and $\phi_{v_k}(t)$, can be derived from $\psi_k(t_1, t_2)$ through the relations:

$$(8)\quad \phi_{x_k^*}(t) = \exp\left\{\int_0^t \frac{\partial\psi_k(t_1,0)/\partial t_2}{\psi_k(t_1,0)}\,dt_1\right\}$$

$$= \exp\left(\int_0^t i\frac{E[w_k e^{it_1x_k}]}{E[e^{it_1x_k}]}\,dt_1\right),$$

$$\phi_{\eta_k}(t) = \frac{\psi_k(t,0)}{\phi_{x_k^*}(t)}.$$

In fact, Susanne M. Schennach (2004a) points out that the equation above shows that, in addition to the full independence between η and $\mathbf{x}^*$, one only needs both u and v to be conditional mean independent of $\mathbf{x}^*$ and η: $E(u\,|\,\mathbf{x}^*,\eta) = 0$ and $E(v\,|\,\mathbf{x}^*,\eta) = 0$.

To summarize the main idea of Li (2002), note that the expressions in (8) represent the marginal characteristic functions of the latent vector of explanatory variables $\mathbf{x}^*$ and the observation errors η in terms of the joint characteristic function of the observable mismeasured variables. They can in turn be used to obtain a complete description of the joint distribution of the unobservable variables.

The conditional distribution for the random vector $\mathbf{x}^*$ given the vectors of observable mismeasured variables $f_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*|\mathbf{x})$ can be written as

$$f_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*|\mathbf{x}) = \frac{f_{\mathbf{x}^*}(\mathbf{x}^*)\prod_{k=1}^K f_{\eta_k}(x_k - x_k^*)}{f_{\mathbf{x}}(\mathbf{x})}.$$

Each component on the right hand side can be estimated as follows. First, the marginal densities of the observable mismeasured variables $f_{\mathbf{x}}(\mathbf{x})$ can be obtained by inverting the Fourier transform of the joint characteristic function of the components of observed variables $\mathbf{x}$. Next, $f_{\mathbf{x}^*}(\mathbf{x}^*)$ can be determined from applying the inverse Fourier transformation to the joint characteristic function of the components of the latent explanatory variable $\mathbf{x}^*$ recovered using knowledge of $\phi_{\eta_k}(t_k)$ from (8): $\phi_{\mathbf{x}^*}(t_1,\cdots,t_K) = \psi_{\mathbf{x}}(t_1,\cdots,t_K)/\prod_{k=1}^K \phi_{\eta_k}(t_k)$. In the empirical implementation, given n independent observations of $\mathbf{x}$, the joint characteristic function of the sample $\psi_{\mathbf{x}}(\cdot)$ is estimated by the sample analog of the product of characteristic functions of individual observations: $\hat{\phi}_{\mathbf{x}}(t_1,\cdots,t_K) = (1/n)\sum_{j=1}^n \exp(\sum_{k=1}^K it_kx_{k,j})$.

A technical problem with the empirical characteristic function is that its inverse Fourier transformation cannot be correctly defined unless its support is "trimmed," because the elements in the summation are not directly integrable. The empirical Fourier transformation is defined by restricting the range of integration to a finite set $[-T_n, T_n]$.

$$\hat{f}_{\mathbf{x}}(x_1,\cdots,x_K)$$
$$= \left(\frac{1}{2\pi}\right)^K \int_{-T_n}^{T_n}\cdots\int_{-T_n}^{T_n} e^{-i\sum_{k=1}^K t_kx_k}$$
$$\times\ \hat{\phi}_{\mathbf{x}}(t_1,\cdots,t_K)dt_1,\cdots,dt_K,$$

where T_n is a "trimming" parameter that is closely related to the bandwidth parameter in kernel smoothing methods, and diverges to infinity at an appropriate rate with increasing sample sizes.

By replacing the true characteristic function with its sample analog, the marginal density of the measurement error is estimated using the formula (8), which relates its characteristic function to the characteristic function of the kth component of the mismeasured variable and the characteristic function of the kth component of latent vector $\mathbf{x}^*$.

Finally, the density $\hat{f}_{\eta_k}(\eta_k)$ itself can be estimated from the truncated version of the inverse Fourier transform suggested above. Similarly, the joint characteristic function of the latent variable $\mathbf{x}^*$ is recovered from the characteristic functions of the mismeasured variables and the measurement errors, which can be inverted subject to truncation to obtain the density of the unobservable regressors $\hat{f}_{\mathbf{x}^*}(\cdot)$. The pointwise convergence rate of the estimated density to the true density of the latent regressor can be established under additional assumptions. These assumptions typically restrict the densities to have finite supports and require that the characteristic functions are uniformly bounded by exponential functions and are integrable on the support.

Given the first step nonparametric estimator $\hat{f}_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*\,|\,\mathbf{x})$, a semiparametric nonlinear least-squares estimator $\hat{\beta}$ for β can be obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \int g(\mathbf{x}^*;\beta)\hat{f}_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*\,|\,\mathbf{x}_i)d\,\mathbf{x}^*\right]^2.$$

Li (2002) establishes the uniform convergence (with rate) of the nonparametric estimator $\hat{f}_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*\,|\,\mathbf{x})$ to the true conditional density $f_{\mathbf{x}^*|\mathbf{x}}(\mathbf{x}^*\,|\,\mathbf{x})$, as well as the consistency of $\hat{\beta}$ to the true parameters of interest β. Importantly, Li (2002) has only demonstrated consistency of his method. Inference methods, including bootstrap or other resampling methods, are not yet available under the assumptions in his model.

The method of Li (2002) can be readily extended to other nonlinear EIV models as long as there are repeated measurements available in the sample; see, e.g., Li and Cheng Hsiao (2004) for consistent estimation of likelihood-based nonlinear EIV models. Furthermore, the deconvolution method via repeated measurement can allow fully nonparametric identification and estimation of models with classical measurement errors and unknown error distributions. See, e.g., Li and Quang Vuong (1998), Li, Isabelle Perrigne, and Vuong (2000), Schennach (2004b), and Stephane Bonhomme and Jean-Marc Robin (2010). Deconvolution techniques are also useful in estimating panel data models with a nonparametric random effect specification. See for example Joel L. Horowitz and Marianthi Markatou (1996).

Recently Schennach (2004a) introduced a somewhat different method for a nonlinear regression model with classical measurement errors. She considers the following model:

$$y = \sum_{k=1}^{M}\beta_k h_k(x^*) + \sum_{j=1}^{J}\beta_{j+M}\, r_j + u,$$

where y and r_j, for $j = 1,\cdots,J$, can be observed, and x^* is the unobserved latent variable with two observable measurements x and w: $x = x^* + \eta$, $w = x^* + v$. The measurement errors are η and v, and u is the disturbance. For convenience, set $r_0 = y$ and use r_j, $j = 0,\cdots,J$, to represent all the observed variables. The assumption of independence between v and x^* is retained, but the conditional mean independence assumptions are relaxed:

$$(9)\quad E[u\,|x^*,\eta] = 0,\quad E[\eta\,|\,x^*,v] = 0,$$

$$E[r_j\,|\,x^*,v] = E[r_j\,|\,x^*],\quad \text{for } j = 1,\cdots,J.$$

In the absence of measurement errors, the population least square regression objective function is

$$E\left[y - \sum_{k=1}^{M} \beta_k h_k(x^*) - \sum_{j=1}^{J} \beta_{j+M} r_j\right]^2.$$

Similar to the reasoning in Hausman et al. (1991), the vector of coefficients β can be identified if the second moments, $E[r_j r_{j'}]$ for j and $j' = 0, 1, \cdots, J$, $E[h_k(x^*)h_{k'}(x^*)]$ for k and $k' = 1, \cdots, M$, and $E[r_j h_k(x^*)]$ for $j = 0, 1, \cdots, J$, and $k = 1, \cdots, M$, are known. Since r_j is observable, its second moment $E[r_j r_{j'}]$ can be estimated by its sample counterpart. However, the two moments $E[h_k(x^*)h_{k'}(x^*)]$ and $E[r_j h_k(x^*)]$ depend on the unobservable latent variable x^*. Schennach (2004a) demonstrates that, by making use of the characteristic function approach, the distribution of x^* and therefore these moments can be related to the sample distribution of the two observable measurements of x^*. The latent moments are recovered from the characteristic function of x^* and the joint features of this characteristic function with other observable variables from sample information.

All the moments required above have the form of $E[W\gamma(x^*)]$, where $W = 1$ when $\gamma(x^*)$ is one of $h_k(x^*)h_{k'}(x^*)$, and $W = r_j, j = 0,\ldots,J$ when $\gamma(x^*)$ is one of $h_k(x^*)$. Theorem 1 in Schennach (2004a) recovers the moments $E[W\gamma(x^*)]$ from observable sampling information through the following relation

$$(10)\quad E[W\gamma(x^*)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mu_\gamma(-\chi)\phi_W(\chi)d\chi,$$

where

$$(11)\quad \phi_W(\chi) \equiv E\left[We^{i\chi x^*}\right] = \frac{E[We^{i\chi w}]}{E[e^{i\chi w}]}\exp\left(\int_0^{\chi} i\,\frac{E[xe^{i\zeta w}]}{E[e^{i\zeta w}]}\,d\zeta\right),$$

and $\mu_\gamma(-\chi)$ is the Fourier transformation of $\gamma(x^*)$ $\mu_\gamma(-\chi) = \int e^{-i\chi x^*}\gamma(x^*)dx^*$. Expression (10) where $\phi_W(\chi) \equiv E[We^{i\chi x^*}]$ follows from substituting the formula for the inverse Fourier transformation of $\mu_\gamma(\chi)$ into the expectation $E[W\gamma(x^*)]$ and subsequent change of the order of integration:

$$E[W\gamma(x^*)] = \int E[W|x^*]\gamma(x^*)f(x^*)dx^*$$

$$= \frac{1}{2\pi}\int\left[\int E[W|x^*]e^{-i\chi x^*}f(x^*)dx^*\right]\mu_\gamma(\chi)d\chi$$

$$= \frac{1}{2\pi}\int \phi_W(-\chi)\mu_\gamma(\chi)d\chi$$

$$= \frac{1}{2\pi}\int \phi_W(\chi)\mu_\gamma(-\chi)d\chi.$$

To show the second equality in (11), we can use the observation that

$$E[We^{i\chi x^*}] = \frac{E[We^{i\chi x^*}]}{E[e^{i\chi x^*}]}E[e^{i\chi x^*}].$$

The last term $E[e^{i\chi x^*}]$ follows from the same derivation of (8) as in Rao (1992), where it is noted that assumption (9) is sufficient. Finally to show $E[We^{i\chi x^*}]/E[e^{i\chi x^*}] = E[We^{i\chi w}]/E[e^{i\chi w}]$, consider the right-hand side

$$\frac{E[We^{i\chi w}]}{E[e^{i\chi w}]} = \frac{E[We^{i\chi(x^*+v)}]}{E[e^{i\chi(x^*+v)}]}$$

$$= \frac{E[We^{i\chi x^*}]E[e^{i\chi v}]}{E[e^{i\chi x^*}]E[e^{i\chi v}]} = \frac{E[We^{i\chi x^*}]}{E[e^{i\chi x^*}]}.$$

The second equality above follows from assumption (9) and the independence between x^* and v. This completes the proof for (10) and (11). Note that when $W = 1$, the first term in (11) vanishes and $\phi_W(\chi) = \phi_{x^*}(\chi)$ is just the characteristic function for x^*.

Given sampling information about y, r_j, x, and w, one can form sample analogs of the population expectations in (10) and (11), and use them to form the estimates for $E[r_j r_{j'}]$, $E[h_k(x^*)h_{k'}(x^*)]$ and $E[r_j h_k(x^*)]$, which are then used to compute the least square estimator. The deconvolution based estimation procedure is a generalization of previous research in polynomial and linear models. If $h_k(x^*)$ is a polynomial, as the case considered in Hausman et al. (1991), given the standard assumptions about the distributions under consideration, the moments of interest $E[W\gamma(x^*)]$ can be used to derive the same estimates as in Hausman et al. (1991). Furthermore, in case of a linear model, this approach is equivalent to the linear IV estimation method.

Schennach (2004a) also considered multivariate measurement errors models, which typically require estimating expectations of the form $E[\gamma(\mathbf{x}^*, \mathbf{r}, \beta)]$. In the multivariate setup, the unobservable variable $\mathbf{x}^*$ is a $K \times 1$ random vector and $\mathbf{x}$ and $\mathbf{w}$ are the corresponding repeated measurements for $\mathbf{x}^*$, such that $\mathbf{x} = \mathbf{x}^* + \eta$ and $\mathbf{w} = \mathbf{x}^* + \mathbf{v}$. To disentangle the characteristic function of the latent vector $\mathbf{x}^*$ we still need to make certain conditional mean independence assumptions between $\mathbf{x}^*$ and $(\mathbf{x}, \mathbf{w})$. The error in $x^{(k)}$, the kth element of the first mismeasured variable $\mathbf{x}$, is assumed to be mean independence of $x^{*(k)}$ and the components of the vector of measurement errors from the second mismeasured variable: $E[\eta^{(k)} \mid x^{*(k)}, v^{(k)}] = 0$. Complete statistical independence is also assumed between the components in the vector of measurement errors in the second mismeasured vector $v^{(k')}$ and the latent vector of explanatory variables $\mathbf{x}^*$, as well as between the observable vector of explanatory variables $\mathbf{r}$ and $v^{(k')}$ for all components $k' = 1,\dots,K$ and $k' \neq k$.

The basic idea of Schennach (2004a) relies on extending Kotlarski's identification result of the density of $\mathbf{x}^*$ through its characteristic function to the identification of the joint density of $f(\mathbf{x}^*, \mathbf{r})$. In particular, the joint characteristic function of $\mathbf{x}^*$ and $\mathbf{r}$ can be identified by

$$(12)\quad \phi_b(\chi, \omega) = E e^{i\chi\mathbf{x}^*} e^{i\omega\mathbf{r}} = E[e^{i\omega\mathbf{r}} e^{i\chi\mathbf{w}}]\left(\prod_{k=1}^{K} E[e^{i\chi_k w^{(k)}}]\right)^{-1} \times \prod_{k=1}^{K} \exp\left(\int_0^{\chi_k} \frac{iE[x^{(k)} e^{i\zeta_k w^{(k)}}]}{E[e^{i\zeta_k w^{(k)}}]} d\zeta_k\right).$$

$\phi_b(\chi, \omega)$ can then be inverted to form the joint density of $\mathbf{x}^*$ and $\mathbf{r}$, and to evaluate the moment condition $E[\gamma(\mathbf{x}^*, \mathbf{r}, \beta)]$ at every possible value of the parameter β. See Schennach (2004a) section 3.3 for details about the asymptotic properties of the estimator.

Schennach (2004a) applies the deconvolution technique to analyze Engel curves of households using data from the Consumer Expenditure Survey. The Engel curve describes the dependence of the proportion of income spent on a certain category of goods on the total expenditure. The author assumes that the total expenditure is reported with error. To reduce the bias in the estimates due to the observational error, the author uses two alternative estimates of the total expenditure. The first estimate is the expenditure reported for the household in the current quarter, while the second estimate is the expenditure reported in the next quarter. The author compares the estimates obtained using the characteristic function approach and the standard feasible generalized least squares estimates. Her estimates show that the feasible generalized least squares estimated elasticities of expenditure on groups of goods with respect to the total expenditure are lower than the elasticities obtained using the deconvolution technique that she provides. This suggests that the method of the author may correct a downward bias in the estimates of income elasticity of consumption that arises from the errors in the observed total expenditure.

3.3 *Nonparametric EIV Models with Generalized Double Measurements*

Carroll et al. (2004) and Fan and Truong (1993) consider a general nonlinear or nonparametric regression model with a mismeasured regressor and a generalized second measurement. In their model the dependent variable y is a function of the latent true regressor x^* and a vector of observed covariates r, and x^* is mismeasured as x. There is also a generalized second measurement w available for the mismeasured regressor x, which follows a varying-coefficient model that is linear in x^* with coefficients being smooth functions of r.

A general specification of the model with covariates r is given by: $y = g(x^*, r) + u$, $E(u|r) = 0, x = x^* + \eta, E(\eta|r) = 0, w = \alpha_0(r) + \alpha_1(r)x^* + \zeta$, $E(\zeta|r) = 0$, $\alpha_1 \neq 0$. The identifying assumptions are that conditional on r, (η, u, ζ) are jointly independent of x^* and are independent of each other, and that $\rho_k(r) = \text{cov}\{[x^* - E(x^*|r)]^k, g(x^*, r)\,|\,r\} \neq 0$ for some positive integer k. If $\text{cov}\{x^*, g(x^*, r)\,|\,r\} \neq 0$ then it suffices to assume that (x^*, u, η, ζ) are mutually uncorrelated conditional on r. These generalizations are more realistic in practice because the magnitude and measurement of w can be very different from those of x.

It is sufficient to identify and estimate $(\alpha_0(r), \alpha_1(r))$. Once these coefficient functions are identified and estimated, one can define $\tilde{w} \equiv (w - \alpha_0(r))/\alpha_1(r)$ so that $\tilde{w} = x^* + v$ with $v \equiv \zeta/\alpha_1(r)$. the framework of section 3.2 can then be applied without modifications with $\tilde{w}$ replacing w.

Under the condition that (η, u, ζ) has mean zero conditional on r, we have $\alpha_0(r) = E[w|r] - \alpha_1(r)\ E[x|r]$. It suffices to identify $\alpha_1(r)$. If cov $\{x^*,\ g(x^*,\ r)\,|\,r\} \neq 0$ then $\alpha_1(r) = \text{cov}(w, y|r)/\text{cov}(x, y|r)$. However, the assumption $\text{cov}\{x^*, g(x^*, r)|r\} \neq 0$ might fail for some classes of functions $g(x^*, r)$. It can be weakened using higher order moments by assuming the existence of some positive integer k such that (for simplicity we consider a simple case of no covariates r)

$$\begin{aligned}\rho_k &= \text{cov}\{g(x^*), [x^* - E(x^*)]^k\}\\ &= \text{cov}\{y, [x - E(x)]^k\} \neq 0.\end{aligned}$$

Empirically, the estimate of k is determined as the first number for which the hypothesis of $\rho_k = 0$ is rejected, or if the null is never rejected, as the number corresponding to the smallest p-value. Once the number k is obtained, the slope coefficient in the "instrument" equation is then identified as

$$\alpha_1 = \text{sign}\{\text{cov}(x, w)\} \left|\frac{\text{cov}[y, [w - E(w)]^k]}{\rho_k}\right|^{1/k}.$$

Once the method of 3.2 is used to identify the distribution of η, the regression function $g(x^*, r)$ can be estimated by nonparametric deconvolution methods as follows. For illustration, consider the simple model without r. The following arguments can be immediately extended to the more general model by conditioning on the variable r.

Let $\phi_\eta(t)$ be the characteristic function of η. Note that $g(x^*) = g(x^*)f(x^*)/f(x^*)$. Under the stated assumptions, both the fourier transformations of $g(x^*)f(x^*)$ and $f(x^*)$ are identified by the data and $\phi_\eta(t)\colon Ee^{itx^*} = Ee^{itx}/\phi_\eta(t)$ and

$$\begin{aligned}\int e^{itx^*} g(x^*) f(x^*) dx^* &= Ee^{itx^*} g(x^*)\\ &= \frac{Ee^{it(x^*+\eta)}(g(x^*) + u)}{\phi_\eta(t)} = \frac{Ee^{itx} y}{\phi_\eta(t)}.\end{aligned}$$

Therefore $g(x^*)$ can be identified through the ratio of inverse fourier transformations of $E e^{itx} y/\phi_\eta(t)$ and $Ee^{itx}/\phi_\eta(t)$.

The nonparametric estimator replaces $E e^{itx} y$ and Ee^{itx} with their empirical analogs by convoluting the empirical distribution of x_i, y_i with a kernel density function with bandwidth h_n. For $\phi_K(\cdot)$ being the Fourier transform of the kernel function $K(\cdot)$, let

$$\hat{E}e^{itx} y \;=\; \phi_K(th_n)\,\frac{1}{n}\sum_{j=1}^{n}\exp(itx_j)y_j,$$

$$\hat{E}e^{itx} \;=\; \phi_K(th_n)\,\frac{1}{n}\sum_{j=1}^{n}\exp(itx_j).$$

Then the nonparametric estimator for $g(x^*)$ can be obtained by replacing $Ee^{itx} y$ and $E e^{itx}$ with $\hat{E}e^{itx} y$ and $\hat{E}e^{itx}$ in the identification formula. Equivalently, it can also be written using the deconvolution kernel as (see Fan and Truong 1993):

$$\hat{g}_n(x^*) \;=\; \frac{\sum_j K_n\!\left(\dfrac{x^*-x_j}{h_n}\right) y_j}{\sum_j K_n\!\left(\dfrac{x^*-x_j}{h_n}\right)},$$

where the transformed kernel function $K_n(\cdot)$ is defined by

$$K_n(x) \;=\; \frac{1}{2\pi}\int \exp(-itx)\,\frac{\phi_K(t)}{\phi_\eta(t/h_n)}\,dt.$$

The authors also suggest that the main regression function can be nonparametrically estimated from the observed variables (y, x, w, r) by applying the previous deconvolution technique conditional on each value of r. As additional methods for estimation, the authors suggest using deconvolution kernels, penalized splines, the SIMEX method, or a Bayesian penalized splines estimator.

It should be noted that the Fan and Truong (1993) nonparametric estimators of $E[ye^{itx}]$ and $E[e^{itx}]$ are general techniques that apply beyond the framework of the current model. By using the convolution of the empirical distribution with a kernel density, such estimators can perform better than those based on the empirical analogs of $E[ye^{itx}]$ and $E[e^{itx}]$. Such estimators could have also been used in sections 3.2 and 3.4.

Carroll et al. (2004) illustrate their estimation procedure using examples from two medical studies. The first study focuses on the analysis of the effect of arsenic exposure on the development of skin, bladder, and lung cancer. The measurement error comes from the fact that physical arsenic exposure (through water) does not necessarily imply that the exposure is biologically active. The application of the suggested method allows the authors to find the effect of the biologically active arsenic exposure on the frequency of cancer incidents. In the other example, the authors study the dependence between cancer incidents and diet. The measurement error comes from the fact that the data on the protein and energy intake come from self-reported food frequency questionnaires, which can record the true food intake with an error. The estimation method suggested in the paper allows the authors to estimate the effect of the structure of the diet on the frequency of related cancer incidents. The instrumental variables are the toenail arsenic measurement in the first example and a measurement of energy intake in the second example.

3.4 *Nonlinear EIV Models with Strong Instrumental Variables*

Although the standard linear instrumental variable assumption is not sufficient to allow for point identification of the parameters in a general nonlinear EIV model, some slightly stronger notions of IVs do imply point identification and consistent estimation. Such instruments are usually referrred to in the statistics, medical, and biology literature as *Berkson models*, in which the latent true variable of interest x^* is predicted (or caused) by the observed random variable z via the causal equation: $x^* = z + \zeta$, where the unobserved random measurement error ζ is assumed to be independent of the observed predictor z. See, e.g., Fuller (1987) and Carroll et al. (2006) for motivations and explanations of the Berkson-error models; and the recent work of Liqun Wang (2004) for identification and estimation of a nonlinear regression model with Berkson measurement errors. Note that this is related to the proxy assumption discussed in section 2.

Although the Berkson-error model might not be a realistic measurement error model to describe many economic data sets, the idea that some observed random variables predict a latent true variable of interest can still be valid in some economics applications, and it makes the same number of assumptions as the classical EIV model. Neither model assumption is nested within the other model assumption. Which assumption is valid is ultimately an empirical question.

Newey (2001) considers the following form of a nonlinear EIV regression model with classical error and a prediction equation: $y = g(x^*, \delta_0) + \varepsilon$, $\quad x = x^* + \eta$, $\quad x^* = \pi_0' z + \sigma_0 \zeta$, where the errors are conditionally mean independent: $E[\varepsilon | z, \zeta] = 0$ and $E[\eta | z, \varepsilon, \zeta] = 0$. The measurement equation $x = x^* + \eta$ contains the classical measurement error η that is statistically independent of x^*. The unobserved prediction error ζ and the "predictor" z in the causal equation $x^* = \pi_0' z + \sigma_0 \zeta$ are assumed to be statistically independent. The vector z is assumed to contain a constant; hence the prediction error ζ is normalized to have zero mean and identity covariance matrix. Apart from the restrictions on the means and variances, no parametric restrictions are imposed on the distributions of the errors. The parameters of interest are $(\delta_0, \pi_0, \sigma_0)$. This model has also been studied in Wang and Hsiao (1995), who proposed similar identification assumptions but a different estimation procedure.

The unknown density of the prediction error ζ, denoted by $f_0(\zeta)$, implies three sets of conditional moment restrictions for conditional expectations of y given z, the product yx given z and the regressor x given z:

$$
\begin{aligned}
(13)\quad E[y\,|\,z] &= E[g(x^*, \delta_0)\,|\,z] \\
&= \int g(\pi_0' z + \sigma_0 \zeta, \delta_0) f_0(\zeta)\, d\zeta, \\
E[y\,x\,|\,z] &= \int [\pi_0' z + \sigma_0 \zeta] \\
&\quad \times\ g(\pi_0' z + \sigma_0 \zeta, \delta_0) f_0(\zeta)\, d\zeta, \\
E[x\,|\,z] &= \pi_0' z.
\end{aligned}
$$

Given that π_0 is identified from $E[x\,|\,z]$, both π_0 and $z^* = E[x\,|\,z]$ can be assumed known for the identification of the remaining parameters in the ensuing discussion.

Newey (2001) suggests a simulated method of moments (SMM) to estimate the parameters of interest $(\delta_0, \pi_0, \sigma_0)$ and the nuisance function $f_0(\zeta)$. Suppose we can simulate from some density $\varphi(\zeta)$. Then represent the density of the error term as: $f(\zeta, \gamma) = P(\zeta, \gamma)\varphi(\zeta)$, where $P(\zeta, \gamma) = \sum_{j=1}^{S} \gamma_j p_j(\zeta)$ for some basis functions $p_j(\cdot)$. The coefficients in the expansion should be chosen so that $f(\zeta, \gamma)$ is a valid density. The coefficient choices need to be normalized to impose restrictions on the first two moments of this density. One possible way of imposing such restrictions is to add them as extra moments into the original system of moments. In the next step, Newey (2001) constructs a system of simulated moments $\hat{\rho}(\alpha)$ for $\alpha = (\delta', \sigma, \gamma')'$ as

$$
\begin{aligned}
\hat{\rho}_i(\alpha, \pi) &= \begin{pmatrix} y_i \\ L x_i\, y_i \end{pmatrix} \\
&\quad - \frac{1}{S} \sum_{s=1}^{S} \begin{pmatrix} g(\pi' z_i + \sigma \zeta_{is}, \delta) \\ L(\pi' z_i + \sigma \zeta_{is}) g(\pi' z_i + \sigma \zeta_{is}, \delta) \end{pmatrix} \\
&\quad \times\ P(\zeta_{is}, \gamma)
\end{aligned}
$$

together with the moment condition $x_i - z_i'\pi$ for π. In the above, L is the matrix selecting the regressors containing the measurement error. Each ζ_{is} is i.i.d. random variables simulated from the density $\varphi(\zeta)$, and S is the total number of simulations.

This system of moments can be used to form a method of moments estimator if the density $f(\zeta)$ takes a parametric form. If $\hat{A}(z_i)$ is a vector of instruments for the observation i, then the sample moment equations will take the form: $m_n(\alpha) = (1/n) \sum_{i=1}^{n} \hat{A}(z_i)\hat{\rho}_i(\alpha)$. The weighting matrix can be obtained from a preliminary estimate for the unknown parameter vector. The standard GMM procedure then follows. Newey (2001) shows that such a procedure will produce consistent estimates of the parameter vector under a set of regularity conditions. If the density $f(\zeta)$ is unknown, the system of three conditional moment equations (13) and Newey's (2001) estimation procedure fit into the framework studied in Chunrong Ai and Xiaohong Chen (2003), whose results are directly applicable to derive root-n asymptotic normality of $(\delta_0, \pi_0, \sigma_0)$ along with a consistent estimate of the asymptotic variance.

Newey's (2001) estimator is applied to the estimation of Engel curves, which models the dependence of the share on a specific commodity group on the income level. The author assumes that individual income is measured with an error that comes in a multiplicative form, allowing for the logarithm of income to enter as the independent variable. In estimation, the author uses the data from the 1982 Consumer Expenditure Survey, giving the shares of individual expenditure on several commodity groups. The estimation method of the paper is implemented for the assumption of a Gaussian error and for the Hermite polynomial specification for the error density, and is compared with the results of the conventional least squares (LS) and the IV estimators. The estimation results show significant downward biases in the LS and IV estimates, while the suggested SMM estimates are close for both the Gaussian specification and the flexible Hermite specification for the distribution of the error term. For example, the median estimates for the elasticity in the share equation for transportation are 1.44 and 1.06 for LS and IV, but 1.01 and 0.98 for the two SMM estimates. This implies that the suggested method can be an effective tool for correcting measurement errors in nonlinear models. The predictor variables z in this example include a constant, age and age squared for household head and spouse, and dummies for educational attainment, spouse employment, home ownership, industry, occupation, region, and race.

The model in Newey (2001) is related to Hausman et al. (1991) and is extended by Schennach (2007) and Victoria Zinde-Walsh (2007) using generalized delta function techniques to a nonparametric setup where the regression function is nonparametrically specified: $y = g(x^*) + u$, $x = x^* + \eta$, $x^* = m(z) + \zeta$. The imposed assumptions include mean independence $E[u \mid z, \zeta] = 0$, $E[\eta \mid z, \zeta, u] = 0$, and the statistical independence of ζ from z. The additional normalization assumption is $E[\zeta] = 0$.

4. *Nonlinear EIV Models with Nonclassical Errors*

The recent applied economics literature has questioned the validity of the classical measurement error assumption. For example, it is often the case that data sets rely on individual respondents to provide information. It may be hard to tell whether or not respondents are making up their answers and, more crucially, whether the measurement error is correlated with the latent true variable and some of the other observed variables. Studies by Bound and Krueger (1991), Bound et al. (1994), Bollinger (1998), and Bound, Brown, and Mathiowetz (2001) have all documented evidence that indicates the violation of the classical measurement errors assumption in economic data sets. This section reviews some of the very recent theoretical advances on nonlinear models with nonclassical measurement errors. Identification in these models typically comes from an instrumental variable assumption or from an additional "auxiliary" sample.

The increasing presence of auxiliary samples has extended the research frontier for measurement error models. In the absence of auxiliary information, researchers are forced to make the classical EIV assumptions in order to obtain point identification of the model parameters. The classical EIV assumption is not testable without auxiliary data. When the auxiliary data becomes available, we can not only test for the validity of the classical EIV assumption but can also devise methods that are consistent without the classical EIV assumption.

These models are often concerned with discrete mismeasured variables, in which case mismeasured variables usually take the form of misclassified discrete responses. For example, a unionized worker might be misclassified as one who is not unionized. When the variable of interest and its measurement are both binary, the measurement error can not be independent of the true binary variable. Typically, misclassification introduces a negative correlation, or mean reversion, between the errors and the true values. As a result, methods developed in section 3 will not generate consistent estimates.

4.1 *Misclassification of Discrete Dependent Variables*

Hausman, Jason Abrevaya, and F. M. Scott-Morton (1998) introduce a modified maximum likelihood estimator to estimate consistently the coefficient parameters and the explicit extent of misclassification in a binary choice model with latent variable $y^* : y_i^* = x_i{}'\beta + \varepsilon_i$, where ε_i is independent of x_i. The probability distribution function of $-\varepsilon_i$ is the same for all i and is denoted as F. The unobserved true response is induced by zero threshold crossing of the latent variable: $\tilde{y}_i = 1(y_i^* \geq 0)$. This response is observed with misclassification, where the misclassified indicator is denoted by y_i. Both the probability of misclassification as one ($\Pr(y_i = 1 \mid \tilde{y}_i = 0, x_i)$) and the misclassification probability as zero ($\Pr(y_i = 0 \mid \tilde{y}_i = 1, x_i)$) are assumed to be independent of the covariates x_i, and are denoted α_0 and α_1 .

Given the above model assumptions, the expected value of the observed misclassified dependent variable is $\Pr(y_i = 1 \mid x_i) = \alpha_0 + (1 - \alpha_0 - \alpha_1)F(x_i{}'\beta) \equiv G(x_i{}'\beta)$. Parameters $(\alpha_0, \alpha_1, \beta)$ of the binary response model with misclassification under the parametrically specified distribution of the disturbance in the latent variable can be estimated by nonlinear least squares or by maximum likelihood. Standard parametric tests for the significance of the coefficients α_0 and α_1 can be used to measure the extent of misclassification in the model. The model of this type cannot be estimated as a "classical" linear probability model where $F(x_i'\,b) = x_i'\,b$ because in that case, one cannot separately identify the parameters of the linear index $x_i'\,\beta$ and the factors α_0 and α_1 .

Identification of the parameters also requires a monotonicity condition that $\alpha_0 + \alpha_1 < 1$, which limits the extent of the misclassification error. In addition, the authors impose a standard invertibility condition requiring that the matrix of regressors $E[xx']$ is nonsingular, and that the distribution function $F(\cdot)$ of the disturbance in the latent variable is known.

The estimates of the model parameters can be used to analyze the influence of misclassification on the parameters in the linear index of the latent variable y^*. Specifically, define β_E (α_0, α_1) to be the probability limits of the misspecified maximum likelihood estimates of β when the mismeasured y_i is used in place of the true $\tilde{y}_i$ in the log likelihood function, as a function of the misclassification probabilities α_0 and α_1. Note that $\beta_E(0,0) = \beta$, the coefficient in the model without misclassification. The marginal effects of misclassification on the parameter estimates can be derived as

$$\left|\frac{\partial \beta_E}{\partial \alpha_0}\right|_{\alpha_0=\alpha_1=0} = -\left[E\left(\frac{f(x'\beta)^2}{F(x'\beta)(1-F(x'\beta))}xx'\right)\right]^{-1} \times E\left(\frac{f(x'\beta)}{F(x'\beta)}x\right),$$

$$\left|\frac{\partial \beta_E}{\partial \alpha_1}\right|_{\alpha_0=\alpha_1=0} = \left[E\left(\frac{f(x'\beta)^2}{F(x'\beta)(1-F(x'\beta))}xx'\right)\right]^{-1} \times E\left(\frac{f(x'\beta)}{1-F(x'\beta)}x\right).$$

The degree of inconsistency of the coefficient estimates using estimation procedures that do not take misclassification into account depends on the distribution of the regressors and the disturbance through its hazard function.

The marginal effects on the conditional choice probabilities are obtained by $\frac{\partial \Pr(\tilde{y}=1|x)}{\partial x} = f(x'\beta)\beta$ for the true response probability, and $\frac{\partial \Pr(y=1|x)}{\partial x} = (1-\alpha_0-\alpha_1)\times f(x'\beta)\beta$ for the observed response probability. When both marginal effects are evaluated at the same parameter value β, their difference is increasing with the magnitude of the misclassification probabilities α_0 and α_1. However, it is important to note that if the observed response is used to estimate the parameter without correcting for measurement errors, the parameter estimate will be generally inconsistent for the true parameter value due to misspecification. The marginal effect on the true choice probability calculated at the true parameter value generally has no clear relation with the marginal effect on the observed response probability that is calculated for the misspecified model.

In many cases, the distribution of disturbances F is unknown. The authors develop identification conditions and estimation procedures for a semiparametric model with a flexible distribution of the error in the latent variable. The identification conditions require the strict monotonicity of $F(\cdot)$ and either that $\alpha_0 + \alpha_1 < 1$ or that $E(y|y^*)$ is increasing in y^*. The first condition $\alpha_0 + \alpha_1 < 1$, being a parametric assumption, is stronger than the second one. It is possible to run a specification test proposed in Horowitz and Wolfgang Hardle (1994). If the parametric model is not rejected, it can be used to improve efficiency.

The regression coefficients, however, are only interesting if they can be used to calculate marginal effects. Correcting for the marginal effect for misclassification requires consistent estimation of the misclassification probabilities α_0 and α_1. In principle, these probabilities can be inferred from the asymptotic behavior of the conditional expectation $E[y|x] = G(x'\beta)$. According to the expression for $G(x'\beta)$ in terms of α_0, α_1 and $F(x'\beta)$, $\lim_{z\to-\infty} G(z) = \alpha_0$ and $\lim_{z\to+\infty} G(z) = 1-\alpha_1$.

The estimation procedure proceeds in two stages. In the first stage, they suggest to estimate the coefficient in the linear index β using the maximum rank correlation (MRC) estimator based on Aaron K. Han (1987): $b_{MRC} = \arg\max_b \sum_{i=1}^n Rank(x_i'b)y_i$. The constant term in b_{MRC} can not be identified, and requires a normalization of the index coefficient. Strong consistency and asymptotic normality of b_{MRC} have been proved by, for example, Han (1987) and Robert P. Sherman (1993). The second stage makes use of the first stage estimate of b_{MRC} and the observed dependent variables to obtain an estimate of the response function $G(\cdot)$ via isotonic regression.

The second stage of the semiparametric estimator converges slower than the parametric first stage due to the nature of isotonic regressions that are used to estimate nonparametric functions. Out-of-sample fit of semiparametric estimates can be poor and in general we cannot use them for precise predictions. However, even without the two stage estimator, the data still provides useful bounding information for the misclassification probabilities and the marginal effect. For example, the marginal effect can be represented as: $\frac{\partial \Pr(\tilde{y}=1|x)}{\partial x} = g(x'\beta)\beta(1-\alpha_0-\alpha_1)^{-1}$, where $g(\cdot)$ is the derivative of $G(\cdot)$. The apparent lower bound for the marginal effect is achieved in the absence of misclassification when the marginal effect is equal to $g(x_i'\beta)\beta$. The results in Horowitz and Charles F. Manski (1995) can be used to form upper bounds for the misclassification probabilities. which will in turn provide an upper bound for the estimated marginal effect.

Hausman, Abrevaya, and Scott-Morton (1998) apply their semiparametric technique to study a model of job change using data from the Current Population Survey and the Panel Study of Income Dynamics. Using these two datasets, the authors can evaluate the probabilities of job change over certain periods of time. According to the authors, the questions about job tenure are not always understood by the respondents and, thus, the survey data contain a certain amount of misclassification error connected with the wrong responses of individuals. Using the methodology of the paper, it is possible to correct the bias in the estimates of the probabilities of job change connected with the misclassification errors in the data. As the authors report, the construction of the job tenure variable in a standard way leads to a substantial bias in the estimates, while the methods provided by the authors allow them to correct the bias due to misclassification.

4.2 *Misclassification of Discrete Regressors Using IVs*

Recently, Aprajit Mahajan (2006) studied a nonparametric regression model where one of the true regressors is a binary variable: $E(y\,|\,x^*, z) = g(x^*, z)$. In this model, the variable x^* is binary and z is continuous. The true binary variable x^* is unobserved and the econometrician observes a potentially misreported value x instead of x^*. Mahajan (2006) assumes that, in addition, another nondegenerate random variable v is observed. The variable v takes at least two values v_1 and v_2, and plays the role of an exclusion restriction analogous to a standard instrumental variable. Naturally, it is required that the regression function $g(x^*, z)$ is identified in the absence of measurement error given the knowledge of the population distribution of $\{y, x^*, z\}$. Mahajan (2006) also imposes the following assumptions on the model.

Denote $\eta_0(z) = \Pr(x=1\,|\,x^*=0, z)$, $\eta_1(z) = \Pr(x=0\,|\,x^*=1, z)$ as the probabilities of misclassification. The second assumption restricts the extent of possible misclassification so that the observed signal is not dominated by misclassification noise.

Assumption 4.1 $\eta_0(z) + \eta_1(z) < 1$.

The third assumption declares independence of the observable misclassified variable from the binary variable v conditional on the true variable x^* and the continuous variable z.

Assumption 4.2 *x is independent of v given x^* and z.*

The next assumption requires that the conditional probability of the true regressor actually depends on the binary instrument v. It implies that the instrumental variable v is informative about the unobserved regressor x^* even given the other covariates.

Assumption 4.3 *There exist* $v_1 \neq v_2$ *such that* $\Pr(x^* = 1 | z, v_1) \neq \Pr(x^* = 1 | z, v_2)$.

The next assumption requires that the unobserved regressor x^* is relevant for the conditional expectation under consideration. This assumption is potentially testable because it implies that the expectation of the variable y conditional on observable x and z should be different for $x = 0$ and $x = 1$.

Assumption 4.4 $g(1, z) \neq g(0, z)$.

The last assumption states that the value of the unobserved regressor x^* provides sufficient information about the outcome variable y so that if the value of x^* is known, then the information about the misclassified variable x and the instrument v is redundant.

Assumption 4.5 $f_{y|x^*,x,z,v}(y | x^*, x, z, v) = f_{y|x^*,z}(y | x^*, z)$.

Mahajan (2006) only imposes a weaker version of assumption 4.5: $E[y | x^*, x, z, v] = E[y | x^*, z]$. We impose this stronger version of assumption 4.5 for simplicity.

Mahajan shows that, under assumptions 4.1–4.5 for almost all z on its support, both the regression function $g(x^*, z)$ and the misclassification probabilities as functions of z are identified.

To see this, denote $\eta_2(z, v) = \Pr(x = 1 | z, v)$ and $\eta_2^*(z, v) = \Pr(x^* = 1 | z, v)$. Note that $\eta_2(z, v)$ is observable and the following relations hold:

$$(14)\quad \begin{cases} E(x | z, v) \equiv \eta_2(z, v) \\ \qquad = (1 - \eta_1(z))\eta_2^*(z, v) \\ \qquad\quad + \eta_0(z)(1 - \eta_2^*(z, v)), \\ E(y | z, v) = g(1, z)\eta_2^*(z, v) \\ \qquad\quad + g(0, z)(1 - \eta_2^*(z, v)), \\ E(yx | z, v) = g(1, z)(1 - \eta_1(z)) \\ \qquad\quad \times \eta_2^*(z, v) + g(0, z)\eta_0(z) \\ \qquad\quad \times (1 - \eta_2^*(z, v)). \end{cases}$$

Suppose v takes n_v values. For each z, $\eta_0(z)$, $\eta_1(z)$, $g(0, z)$, $g(1, z)$ and $\eta_2^*(z, v)$ are unknown. There are $4 + n_v$ parameters, and $3n_v$ equations. Therefore as long as $n_v \geq 2$, all the parameters can possibly be identified.

A constructive proof is given in Mahajan (2006) using the above three moment conditions.

The solutions to this system of equations is only unique up to an exchange between $1 - \eta_1(z)$ and $\eta_0(z)$. However, assumption 4.1 rules out one of these two possibilities and allows for point identification; hence the model is identified.

Mahajan (2006) further develops his identification strategy into a nonparametric estimator, which follows closely the identification argument of the model. Specifically, one estimates the system of moments (14) by kernel smoothing and then solves it for the $4 + n_v$ unknown parameters. The appropriate asymptotic behavior of the obtained estimates can be derived under additional conditions. Mahajan (2006) also provides a semiparametric estimator for a single index model. The semiparametric model can be further simplified to a fully parametric model such as a parametric binary choice model. In the parametric case, the additional identification assumptions reduce to the invertability of the matrix of independent variables and a nonzero coefficient for the unobserved binary variable.

Mahajan (2006) also suggests a constructive and simple test for misclassification. The idea of the test is that the instrument v is relevant for the outcome variable y only in the case where there is misclassification. Otherwise the information about the true regressor x^* is sufficient for the conditional expectation of the outcome y. He first proves that both misclassification probabilities are zero $\eta_0(z) = \eta_1(z) = 0$ if and only if the instrument v is not relevant, so that $E(y | x, z, v) = E(y | x, z)$. The test for misclassification can be conducted as a test for the equality of the two conditional expectations, both of which can be estimated nonparametrically. Nonparametric tests based on appropriate orthogonality and moment conditions have been developed in, for example, Pascal Lavergne and Vuong (2000) and Yanqin Fan and Qi Li (1996). Under the null hypothesis, their test statistics converge to a normal distribution with zero mean at a nonparametric rate.

An alternative approach to identification and estimation of a model with a misclassified discrete regressor is considered in the paper by Yingyao Hu (2008). This paper looks at a general problem of identification of the joint density: $f_{y\,|\,x^*,z}(y\,|\,x^*,z)$. Here y is a one-dimensional random variable, x^* is the unobserved discrete regressor, and z is the observed regressor. One can observe a misclassified variable for the unobserved regressor x^*—a discrete variable x and an instrument v. It is assumed that the variables x, x^* and v have a common discrete support $\{1,2,\ldots,k\}$. Analogs of assumptions 4.1 to 4.5 of Mahajan (2006) continue to hold.

Under assumptions 4.2 and 4.5, the author suggests that one can form a system of equations relating the observed distributions to the unobserved distributions and then find the unobserved distributions by matrix operations. This approach generalizes the insight of Mahajan (2006) to the case of multiple values of the misclassified regressor.

To develop the identification and estimation results, define the following matrices:

$$\begin{aligned}
\mathbf{F}_{yx|vz} &= (f_{yx|vz}(y,i\,|\,j,z))_{i,j=1}^{k},\\
\mathbf{F}_{x^*|vz} &= (f_{x^*|vz}(i\,|\,j,z))_{i,j=1}^{k},\\
\mathbf{F}_{x|x^*z} &= (f_{x|x^*z}(i\,|\,j,z))_{i,j=1}^{k},\\
\mathbf{F}_{y|x^*z} &= \operatorname{diag}\{f_{y|x^*z}(y\,|\,i,z),\ i = 1,\ldots,k\},\\
\mathbf{F}_{y|vz} &= (f_{y|vz}(y\,|\,i,z))_{i=1}^{k},\\
\mathbf{F}_{x|vz} &= (f_{x|vz}(x\,|\,i,z))_{i=1}^{k}
\end{aligned}$$

In these definitions, we use index i for columns and index j for rows. Under assumptions 4.2 and 4.5, the conditional distributions are expressed in matrix form as

$$\begin{aligned}
\mathbf{F}_{yx|vz} &= \mathbf{F}_{x^*|vz}\,\mathbf{F}_{y|x^*z}\,\mathbf{F}_{x|x^*z},\\
\mathbf{F}_{x|vz} &= \mathbf{F}_{x^*|vz}\,\mathbf{F}_{x|x^*z}.
\end{aligned} \tag{15}$$

An additional equation comes from the definition of the conditional density:

$$\mathbf{F}_{y|vz} = \mathbf{F}'_{x^*|vz}\,\mathbf{F}_{y|x^*z}\,\mathbf{1}, \tag{16}$$

where $\mathbf{1}$ is a $k \times 1$ vector of ones. To complete the system of equations (15) and (16), the author adds the following assumption, which generalizes assumption 4.4:

Assumption 4.6 *Rank*$(\mathbf{F}_{x^*|v,z}) = k$.

Assuming in addition nonsingularity of the matrix $\mathbf{F}_{x|x^*z}$, we can add a system of equations for the unknown $k(k+1)$ elements of $\mathbf{F}_{y|x^*z}$ and $\mathbf{F}_{x|x^*z}$ for every possible y and z:

$$\begin{aligned}
\mathbf{F}_{x|x^*z}\,\mathbf{F}^{-1}_{x|vz}\,\mathbf{F}_{yx|vz}\,\mathbf{F}^{-1}_{x|x^*z} &= \mathbf{F}_{y|x^*z} \text{ and}\\
\mathbf{F}_{x|x^*z}\,\mathbf{1} &= \mathbf{1}.
\end{aligned}$$

Construct a matrix $\mathbf{A} = \mathbf{F}^{-1}_{x|vz}\,\mathbf{F}_{yx|vz}$ from the matrices of observable distributions. Since the matrix $\mathbf{F}_{y|x^*z}$ is diagonal and is expressed in a "sandwich" form in terms of the matrix $\mathbf{A}$, $\mathbf{F}_{y|x^*z}$ and $\mathbf{A}$ have the same eigenvalues. Even though the matrix $\mathbf{A}$ can be reconstructed from the data, additional assumptions need to be imposed to map its eigenvalues to the elements of the matrix $\mathbf{F}_{y|x^*z}$. The author imposes two additional restrictions. First, there is a function $\gamma(\cdot)$ such that the expectation $E[\gamma(y)\,|\,x^* = i, z] \neq E[\gamma(y)\,|\,x^* = j, z]$ for all $i \neq j$. Second, the conditional distribution $f_{y|x^*z}$ is strictly monotone in x^* for every y and z. Under these additional restrictions, the values of the densities $\mathbf{F}_{y|x^*z}$ can be associated with the ordered eigenvalues of the matrix **A**. Furthermore, the matrix of eigenvectors of the matrix **A** can be associated with the matrix of misclassification probabilities $\mathbf{F}_{x|x^*z}$. Hu and Schennach (2008) extend the identification approach of Hu (2008) to nonclassical measurement error models where the latent variable x^*, and observed variables y, x and instruments v are all continuous ramdon variables.

The estimation strategy suggested in Hu (2008) is suited for a semiparametric specification when the outcome variable is described by a nonlinear parametric regression function: $E(y\,|\,x^*, z) = g(x^*, z; \theta_0)$. After the unknown distribution $f_{x^*|x,z}$ is obtained from the eigenvalue decomposition following the procedure described in the identification proof, the regression function $g(\cdot)$ can be transformed to a function of observable variables y, x, z, and v to form a GMM-type objective function. Given a set of smoothness and uniform boundedness assumptions, the author proves that the estimate of the parameter θ_0 is asymptotically normal with a parametric $\sqrt{n}$ rate of convergence.

As an application, Hu (2008) analyzes the impact of education on women's fertility using data from the Current Population Survey. The distribution of the dependent variable is characterized by a quasi-maximum likelihood estimator (QMLE) based on the Poisson distribution. The author compares the performance of the method developed in his paper with the performance of a naive standard QMLE that does not account for the presence of misclassification errors. The naive QMLE estimates of the semi-elasticity of the number of children with respect to education are biased toward zero. This implies that failure to account for the presence of measurement errors can potentially underevaluate the effects of policy changes. The estimates of the semi-elasticity obtained using the author's correction are almost twice the estimates obtained using the QMLE. A Hausman-type test is used as a specification test to verify the presence of the measurement error in the data. The test rejects the hypothesis of the absence of the measurement errors. The instruments are parents' education and the women's number of children.

In a related paper, Lewbel (2007) considers a model of the average treatment effect when the treatment is observed with an error. This model is closely related to Mahajan (2006) in that it makes use of an instrumental variable condition. The key difference is that Lewbel (2007) assumes that the instrument is conditionally independent of the conditional average effect, while Mahajan (2006) assumes that the instrument is conditionally independent of the outcome itself. Because the conditional average effect, which is the difference between two outcome statuses, is itself a function of two unknown misclassification probabilities, Lewbel (2007) requires three values of the instrumental variable for identification, while Mahajan (2006) only requires two different values of the instrumental variable.

Specifically, the object of interest is the conditional average effect on y of changing the true treatment status from 0 to 1:

$$\text{(17)} \qquad \tau^*(z, v) \;=\; E(y\,|\,z, v, x^* \;=\; 1) \;-\; E(y\,|\,z, v, x^* \;=\; 0).$$

The true binary regressor x^* is mismeasured as x. Y is the observed treatment outcome, and (z, v) is the set of instruments. In addition, if we define y_0 to be the variable corresponding to the treatment outcome when $x^* = 0$ and y_1 as the outcome when $x^* = 1$, then the conditional average treatment effect is defined as

$$\text{(18)} \qquad \tilde{\tau}(z, v) \;=\; E[y_0 - y_1\,|\,z, v].$$

Under the unconfoundedness assumption, the conditional average effect on y is equivalent to the conditional average treatment effect. The identification results of Lewbel

(2007) apply to the conditional average effect of (17). With the unconfoundedness assumption, they also apply to the conditional average treatment effect (18).

Analogs of assumptions 4.5 and 4.1 are needed for identification:

$$(19)\quad E(y\,|\,z,v,x^*,x) \;=\; E(y\,|\,z,v,x^*).$$

$$\Pr(x \;=\; 0\,|\,z,v,x^* \;=\; 1) + \Pr(x \;=\; 1\,|\,z,v,x^* \;=\; 0) \;<\; 1.$$

It is also assumed that the treatment probability is positive but not all of the outcomes are treated: $0 < r^*(z,v) = E(x^*\,|\,z,v) < 1$.

Under these assumptions, the author proves that if $\tau(z,v)$ is the observed treatment effect (estimated from the mismeasured treatment dummy), then the relation between the true treatment effect and the observed treatment effect is given by a function that depends on the two unknown misclassification probabilities and the observed treatment probability $r(z,v)$:

$$(20)\quad \tau^*(z,v) = \tau(z,v)h(r(z,v), \Pr(x=0\,|\,z,v,x^*=1), \Pr(x=1\,|\,z,v,x^*=0)).$$

To jointly identify $\tau^*(z,v)$ and the misclassification probabilities, the author imposes two additional assumptions.

The first assumption requires that for some subset of the support of (z,v), we can fix z and variation in v does not lead to changes in the conditional misclassification probabilities and the true treatment effect, but changes the probability of the true treatment dummy. Formally, there exists a known set $\mathcal{A}$ in the support of z and v such that for all $((z,v)\,(z,v')) \in \mathcal{A}$ where $v' \neq v$, $r^*(z,v), \neq r^*(z,v')$, but: $\Pr(x=1\,|\,z,v,x^*=0) = \Pr(x=1\,|\,z,v',x^*=0)$, $\Pr(x=0\,|\,z,v,x^*=1) = \Pr(x=0\,|\,z,v',x^*=1)$ and $\tau^*(z,v) = \tau^*(z,v')$.

The last identifying assumption imposes a testable "sufficient variation" restriction, which assumes that it is possible to find three elements in the support of (z,v) with v_0, v_1, v_2, and the same component z, such that

$$\left(\frac{\tau(v_0,z)}{r(v_1,z)} - \frac{\tau(v_1,z)}{r(v_0,z)}\right)\left(\frac{\tau(v_0,z)}{1-r(v_2,z)} - \frac{\tau(v_2,z)}{1-r(v_0,z)}\right)$$

$$\neq \left(\frac{\tau(v_0,z)}{r(v_2,z)} - \frac{\tau(v_2,z)}{r(v_0,z)}\right)\left(\frac{\tau(v_0,z)}{1-r(v_1,z)} - \frac{\tau(v_1,z)}{1-r(v_0,z)}\right).$$

Under these assumptions, the true treatment effect $\tau^*(z,v)$, the misclassification probabilities $P(x\,|\,z,v,x^*)$ and the probability of treatment $r^*(z,v)$ are all identified.

Intuitively, the three "unknown" variables in (20), the left-hand side $\tau^*(z,v)$ and the two misclassification probabilities do not depend on v. One needs three equations to identify three unknown parameters. Therefore three different values of v are needed. The "sufficient variation" condition is a "rank condition" for the system of nonlinear equations formed by (20) at three different values of v.

Lewbel (2007) suggests a GMM estimation method when the support of the instrument v is discrete with K elements, $\{v_k, k=1,\ldots,K\}$. The estimation is based on two moments. The first moment equation expresses the unconditional probability of the observed treatment dummy in terms of the probabilities of mismeasurement. The second moment equation makes use of the relationship between the observed and the true treatment effect.

The GMM procedure solves for the unknown functions assuming a parametric form for the unknown probability distributions and a semiparametric specification for the distribution of the covariates z.

Lewbel (2007) then applies his identification and estimation procedure to study the effect of having a college degree on earnings. Data from the National Longitudinal Survey of the high school class of 1972 provides information about wages. Information about completion of a college degree comes from transcripts in the Post-secondary Education Transcript Survey, and can be misreported. A discrete support instrument comes from the rank data about the distance from the respondent's high school to the closest four-year college. Experience and demographic variables are used as additional covariates. To simplify the analysis, the author suggests a parametric specification for the probability of misreporting, the probability of the true binary regressor (indicating the college degree), and the treatment effect, which is assumed to depend linearly on covariates. Then the parameters of interest are estimated by GMM. The author finds that misclassification introduces a significant downward bias on the estimates: "naive" estimation gives an impact of 11 percent from the college degree relative to a high school diploma, while the GMM method estimates an impact of 38 percent. This is broadly consistent with the estimate of about 25 percent reported in Kane, Rouse, and Staiger (1999) after combining transcript and self report data.

4.3 *Models with Auxiliary Data and Discrete Mismeasured Regressors*

Recently Chen and Hu (2006) study general nonlinear models with nonclassical measurement errors using two samples, both of which contain measurement errors. Neither sample contains an accurate observation of the truth or a valid strong instrumental variable. We illustrate their identification strategy using a special case in which the key variables in the model are dichotomous. Suppose that we are interested in the effect of the true college education level X^* on the labor supply Y. Additional conditioning covariates are the marital status R^u and the gender R^v. This effect would be identified if we could identify the joint density $f_{X^*,R^u,R^v,Y}$.

Assume that X^*, R^u, and R^v are all dichotomous. The true education level X^* is unobserved and is misreported as X. In both the primary sample and the auxiliary sample, (R^u, R^v) are accurately measured and observed, but Y is only observed in the primary sample. The primary sample is a random sample from (X, R^u, R^v, Y). The auxiliary sample is a nonrandom sample of observations of (X_a, R^u_a, R^v_a), in which the observed X_a is a mismeasurement of true latent true education level X^*_a. Chen and Hu (2006) require the following key assumption to nonparametrically identify $f_{X^*,R^u,R^v,Y}$: the measurement error in X is independent of all other variables in the model conditional on the true value X^*, *i.e.*, $f_{X|X^*,R^u,R^v,Y} = f_{X|X^*}$. Under this assumption, the probability distribution of the observables equals

$$(21)\quad f_{X,R^u,R^v,Y}(x,u,v,y) = \sum_{x^*=0,1} f_{X|X^*}(x|x^*) f_{X^*,R^u,R^v,Y}(x^*,u,v,y)$$

for all x, u, v, y.

Define the matrix representation of $f_{X|X^*}$ as follows:

$$L_{X|X^*} = \begin{pmatrix} f_{X|X^*}(0|0) & f_{X|X^*}(0|1) \\ f_{X|X^*}(1|0) & f_{X|X^*}(1|1) \end{pmatrix},$$

Equation (21) then implies for all u, v, y

$$(22)\quad \begin{pmatrix} f_{X,R^u,R^v,Y}(0,u,v,y) \\ f_{X,R^u,R^v,Y}(1,u,v,y) \end{pmatrix} = L_{X|X^*} \times \begin{pmatrix} f_{X^*,R^u,R^v,Y}(0,u,v,y) \\ f_{X^*,R^u,R^v,Y}(1,u,v,y) \end{pmatrix}.$$

Equation (22) implies that the density $f_{X^*,R^u,R^v,Y}$ would be identified provided that

$L_{X|X^*}$ would be identifiable and invertible. Moreover, equation (21) implies, for the subsamples of either males ($R^v = 1$) or females ($R^v = 0$)

$$(23) \quad f_{X,R^u|R^v=j}(x,u) = \sum_{x^*=0,1} f_{X|X^*,R^u,R^v=j}(x|x^*,u) \times f_{R^u|X^*,R^v=j}(u|x^*) f_{X^*|R^v=j}(x^*).$$
$$= \sum_{x^*=0,1} f_{X|X^*}(x|x^*) f_{R^u|X^*,R^v=j}(u|x^*) f_{X^*|R^v=j}(x^*),$$

in which $f_{X,R^u|R^v=j}(x,u) \equiv f_{X,R^u|R^v}(x,u|j)$ and $j = 0, 1$.

Similar assumptions are made for the auxiliary sample. The two samples are linked by a stability assumption that the distribution of the marital status conditional on the true education level and gender is the same in the two samples, i.e., $f_{R^u_a|X^*_a,R^v_a=j}(u|x^*) = f_{R^u|X^*,R^v=j}(u|x^*)$ *for all* u, j, x^*. Intuitively, the stability assumption allows for this distribution to be canceled out in two functional relations that represent respectively the primary sample and the auxiliary sample. In the combined equation, $L_{X|X^*}$ can then be viewed as an identifiable eigenvector.

Under this assumption, for both the subsamples of males ($R^v_a = 1$) and females ($R^v_a = 0$):

$$(24) \quad f_{X_a,R^u_a|R^v_a=j}(x,u) = \sum_{x^*=0,1} f_{X_a|X^*_a}(x|x^*) f_{R^u|X^*,R^v=j}(u|x^*) f_{X^*_a|R^v_a=j}(x^*).$$

Define the matrix representations of relevant densities for the subsamples of males ($R^v = 1$) and of females ($R^v = 0$) in the primary sample as follows: for $j = 0,1$,

$$L_{X,R^u|R^v=j} = \begin{pmatrix} f_{X,R^u|R^v=j}(0,0) & f_{X,R^u|R^v=j}(0,1) \\ f_{X,R^u|R^v=j}(1,0) & f_{X,R^u|R^v=j}(1,1) \end{pmatrix}$$

$$L_{R^u|X^*,R^v=j} = \begin{pmatrix} f_{R^u|X^*,R^v=j}(0|0) & f_{R^u|X^*,R^v=j}(1|0) \\ f_{R^u|X^*,R^v=j}(0|1) & f_{R^u|X^*,R^v=j}(1|1) \end{pmatrix}$$

$$L_{X^*|R^v=j} = \begin{pmatrix} f_{X^*|R^v=j}(0) & 0 \\ 0 & f_{X^*|R^v=j}(1) \end{pmatrix}.$$

Similarly, define the matrix representations $L_{X_a,R^u_a|R^v_a=j}$, $L_{X_a|X^*_a}$, $L_{R^u_a|X^*_a,R^v_a=j}$, and $L_{X^*_a|R^v_a=j}$ of the corresponding densities $f_{X_a,R^u_a|R^v_a=j}$, $f_{X_a|X^*_a}$, $f_{R^u_a|X^*_a,R^v_a=j}$ and $f_{X^*_a|R^v_a=j}$ in the auxiliary sample. Note that equation (23) implies for $j = 0, 1$,

$$(25) \quad L_{X,R^u|R^v=j} = L_{X|X^*} L_{X^*|R^v=j} L_{R^u|X^*,R^v=j}.$$

Similarly, equation (24) implies that

$$(26) \quad L_{X_a,R^u_a|R^v_a=j} = L_{X_a|X^*_a} L_{X^*_a|R^v_a=j} L_{R^u|X^*,R^v=j}.$$

One can then eliminate $L_{R^u|X^*,R^v=j}$, to have for $j = 0, 1$

$$L_{X,R^u|R^v=j} L^{-1}_{X,R^u_a|R^v_a=j} = L_{X|X^*} L_{X^*|R^v=j} L^{-1}_{X^*_a|R^v_a=j} L^{-1}_{X_a|X^*_a}.$$

Since this equation holds for $j = 0, 1$, one may then eliminate $L_{X|X^*}$, to have

$$(27) \quad L_{X,X} \equiv (L_{X,R^u|R^v=1} L^{-1}_{X_a,R^u_a|R^v_a=1}) \times (L_{X,R^u|R^v=0} L^{-1}_{X_a,R^u_a|R^v_a=0})^{-1}$$
$$= L_{X|X^*}(L_{X^*|R^v=1} L^{-1}_{X^*_a|R^v_a=1} \times L_{X^*|R^v=0} L^{-1}_{X^*_a|R^v_a=0}) \times L^{-1}_{X|X^*}.$$

Notice that the matrix $k_{X_a^*} = L_{X^*|R^v=1} L_{X_a^*|R_a^v=1}^{-1} L_{X_a^*|R_a^v=0} L_{X^*|R^v=0}^{-1}$ is diagonal because $L_{X^*|R^v=j}$ and $L_{X_a^*|R_a^v=j}$ are diagonal matrices. Equation (27) provides an eigenvalue–eigenvector decomposition of an observed matrix $L_{X,X}$ on the left-hand side.

Therefore for an observed $L_{X,X}$, one may use an eigenvalue–eigenvector decomposition to identify $k_{X_a^*}$ and $L_{X|X^*}$ if $k_{X_a^*}(0) \neq k_{X_a^*}(1)$; i.e., the eigenvalues are distinctive. This assumption requires that the distribution of the latent education level of males or females in the primary sample is different from that in the auxiliary sample, and that the distribution of the latent education level of males is different from that of females in one of the two samples. Because each column in $L_{X_a|X_a^*}$ is a conditional density, each eigenvector is automatically normalized.

To resolve the ambiguity in the indexing of the eigenvalues and eigenvectors, the authors make a normalization assumption that people with (or without) college education in the auxiliary sample are more likely to report that they have (or do not have) college education; i.e., $f_{X_a|X_a^*}(x^*|x^*) > 0.5$ *for* $x^* = 0, 1$. This assumption also implies the invertibility of $L_{X_a|X_a^*}$, and pins down the index x_1^* as follows:

$$x_1^* = \begin{Bmatrix} 0 & \text{if} f_{X_a|X_a^*}(0|x_1^*) > 0.5 \\ 1 & \text{if} f_{X_a|X_a^*}(1|x_1^*) > 0.5 \end{Bmatrix}.$$

Finally, the density of interest $f_{X^*,R^u,R^v,Y}$ is identified from equation (22).

4.4 *Models with Auxiliary Data and Possibly Continuous Regressors*

In order to allow for possibly continuous mismeasured variables and nonclassical measurement errors, Chen, Hong, and Tamer (2005) and Chen, Hong, and Alessandro Tarozzi (2008) also make use of an auxiliary data set but employ a different identification assumption. They are interested in obtaining consistent estimates of the parameters β in the moment condition $E[m(Y^*;\beta)] = 0$. An auxiliary data set is used to recover the correlation between the measurement errors and the underlying true variables by providing information about the conditional distribution of the measurement errors given the observed reported variables or proxy variables. In contrast to the discrete case, their assumption essentially requires that all variables be present in the auxiliary data set. In their model, the auxiliary data set is a subset of the primary data, indicated by a dummy variable $D = 0$, which contains both the reported variable Y and the validated true variable Y^*. Y^* is not observed in the rest of the primary data set ($D = 1$), which is not validated. They assume that the conditional distribution of the true variables given the reported variables can be recovered from the auxiliary data set:

Assumption 4.7 $Y^* \perp D | Y$.

Assumption 4.7 allows the auxiliary data set to be collected using a **stratified sampling** design where a *nonrandom response based subsample* of the primary data is validated. In a typical example of this stratified sampling design, we first oversample a certain subpopulation of the mismeasured variables Y, and then validate the true variables Y^* corresponding to this nonrandom stratified subsample of Y. It is very natural and sensible to oversample a subpopulation of the primary data set where more severe measurement error is suspected to be present. Assumption 4.7 is valid as long as, in this sampling procedure of the auxiliary data set, the sampling scheme of Y in the auxiliary data is based only on the information available in the distribution of the primary data set $\{Y\}$.

Under assumption 4.7, an application of the law of iterated expectations gives

$$E[m(Y^*;\beta)] = \int g(Y;\beta) f(Y) dY,$$

where $g(Y;\beta) = E[m(Y^*;\beta) \mid Y, D = 0]$.

This suggests a semiparametric GMM estimator for the parameter β. For each value of β in the parameter space, the conditional expectation function $g(Y;\beta)$ can be nonparametrically estimated using the auxiliary data set where $D = 0$.

Chen, Hong, and Tamer (2005) use the method of sieves to implement this nonparametric regression. Let n denote the size of the entire primary dataset and let n_a denote the size of the auxiliary data set where $D = 0$. For $\lambda = n/n_a$, they require that $0 < \lambda < \infty$. Let $\{q_l(Y), l = 1, 2, \ldots\}$ denote a sequence of known basis functions that can approximate any square-measurable function of Y arbitrarily well. Also let $q^{k(n_a)}(Y) = (q_1(Y), \ldots, q_{k(n_a)}(Y))'$ and $Q_a = (q^{k(n_a)}(Y_{a1}), \ldots, q^{k(n_a)}(Y_{an_a}))'$ for some integer $k(n_a)$, with $k(n_a) \to \infty$ and $k(n_a)/n \to 0$ when $n \to \infty$. In the above, Y_{aj} denotes the jth observation in the auxiliary sample. Then for each given β, the first step nonparametric estimation can be defined as,

$$\hat{g}(Y;\beta) = \sum_{j=1}^{n_a} m(Y^*_{aj};\beta) q^{k(n_a)}(Y_{aj})(Q_a' Q_a)^{-1} q^{k(n_a)}(Y).$$

A GMM estimator for β_0 can then be defined as the minimizer of a quadratic norm of $\hat{g}(Y;\beta)$ using an appropriate weighting matrix $\hat{W}$. Chen, Hong, and Tarozzi (2008) show that a proper choice of $\hat{W}$ achieves the semiparametric efficiency bound for the estimation of β. They called this estimator the *conditional expectation projection GMM* estimator.

A special case of assumption 4.7 is when the auxiliary data is generated from the same population as the primary data, where a full independence assumption is satisfied:

Assumption 4.8 $Y, Y^* \perp D$.

This case is often referred to as a (true) **validation sample**. Semiparametric estimators that make use of a validation sample include Carroll and M. P. Wand (1991), Jungsywan H. Sepanski and Carroll (1993), Lung-fei Lee and Sepanski (1995), and the recent work of Paul J. Devereux and Gautam Tripathi (2009). Interestingly, in the case of a validation sample, Lee and Sepanski (1995) suggested that the nonparametric estimation of the conditional expectation function $g(Y;\beta)$ can be replaced by a finite dimensional linear projection $h(Y;\beta)$ into a fixed set of functions of Y. In other words, instead of requiring that $k(n_a) \to \infty$ and $k(n_a)/n \to 0$, we can hold $k(n_a)$ to be a fixed constant in the above least square regression for $\hat{g}(Y;\beta)$. Lee and Sepanski (1995) show that this still produces a consistent and asymptotically normal estimator for β as long as the auxiliary sample is also a validation sample that satisfies the full independence assumption 4.8. However, if the auxiliary sample satisfies only the conditional independence assumption 4.7 but not the full independence assumption 4.8, then it is necessary to require $k(n_a) \to \infty$ to obtain consistency. Furthermore, even in the case of a validation sample, requiring $k(n_a) \to \infty$ typically leads to a more efficient estimator for β than a constant $k(n_a)$.

An alternative consistent estimator that is valid under assumption 4.7 is based on the inverse probability weighting principle, which provides an equivalent representation of the moment condition $Em(y^*;\beta)$. Define $p(Y) = p(D = 1 \mid Y)$ and $p = P(D = 1)$. Assumption 4.7 implies that $f(Y^* \mid Y, D = 0) = f(Y^* \mid Y)$ and, therefore,

$$Em(y^*;\beta) = E\left[m(Y^*;\beta_0)\frac{1-p}{1-p(Y)} \,\middle|\, D = 0\right].$$

To see this, note that,

$$
\begin{aligned}
&E\left[m(Y^*;\beta_0)\,\frac{1-p}{1-p(Y)}\,\middle|\, D=0\right].\\
&= \int m(Y^*;\beta_0)\,\frac{1-p}{1-p(Y)}\\
&\quad\times \frac{f(Y)(1-p(Y))\,f(Y^*\,|\,Y,D\,=\,0)}{1-p}\,dY^*dY\\
&= \int m(Y^*;\beta_0)\,f(Y^*\,|\,Y)\,f(Y)\,dY^*dY\\
&\quad = E\,m(y^*;\beta).
\end{aligned}
$$

This equivalent reformulation of the moment condition $E\,m(Y^*;\beta)$ suggests a two-step *inverse probability weighting GMM* estimation procedure. In the first step, one typically obtains a parametric or nonparametric estimate of the so-called propensity score $\hat{p}(Y)$ using for example a logistic binary choice model with a flexible functional form. In the second step, a sample analog of the reweighted moment conditions is computed using the auxiliary data set: $\hat{g}(\beta) = \frac{1}{n_a}\sum_{j=1}^{n_a} m(Y_j^*;\beta)\,\frac{1}{1-\hat{p}(Y_j)}$. This is then used to form a quadratic norm to provide a GMM estimator: $\hat{\beta} = \arg\min_\beta \hat{g}(\beta)'\,W_n\,\hat{g}(\beta)$. When the propensity score is parametrically specified, it needs to be correctly specified in order for the estimator to be consistent.

Interestingly, an analog of the conditional independence assumption 4.7 is also rooted in the program evaluation literature and is typically referred to as the assumption of unconfoundedness, or selection based on observables. Semiparametric efficiency results for the mean treatment effect parameters for nonlinear GMM models have been developed by, among others, James M. Robins, Steven D. Mark, and Newey (1992), Jinyong Hahn (1998), and Keisuke Hirano, Guido W. Imbens, and Ridder (2003). Many of the results presented here generalize these results for the mean treatment effect parameters to nonlinear GMM models.

An example of a GMM-based estimation procedure that achieves the semiparametric efficiency bound can be found in Chen, Hong, and Tarozzi (2008). In particular, Chen, Hong, and Tarozzi (2008) show that both a semiparametric conditional expectation projection estimator and a semiparametric propensity score estimator based on a sieve nonparametric first stage regression achieve the semiparametric efficiency bound.

Other recent papers that develop estimation methods using combined samples include Oliver Linton and Yoon-Jae Whang (2002), Devereux and Tripathi (2009), Hidehiko Ichimura and Elena Martinez-Sanchis (2005), and Hu and Ridder (forthcoming).

5. *Contaminated and Corrupted Data*

The previous sections rely on strong assumptions regarding the structure of the data and the distribution of measurement errors that may not always be appropriate. While the model is not identified without making these strong assumptions, it is still possible to construct identified sets of parameters under weaker conditions. A recent literature of partial inference aims at providing set estimators that are robust when every observation in the sample can be contaminated or corrupted with a certain probability.

Under a weak assumption that an upper bound can be put on the probability of data error, Horowitz and Manski (1995) provide consistent and sharp bounds on the latent true distribution function when the data is subject to contamination. They consider the problem of estimating the marginal distribution of the unobserved true random variable y^*, which is sometimes misreported due to contamination by a measurement error. In their model, the observed variable $y \equiv y^*(1-d) + \tilde{y}d$, where $\tilde{y}$ is the erroneous response, y^* is the true response and d is a binary variable indicating whether the observation is contaminated or not: $d \in \{0, 1\}$. If

$d = 0$, the observation of the response variable $y = y^*$ is free of error. If $d = 1$, $y = \tilde{y}$ is a contaminated observation. It is assumed that the variables y^* and $\tilde{y}$ have a common support Y.

Contaminated and corrupted data can be viewed as a special case of continuous variables with nonclassical errors since $y \equiv y^*(1-d) + \tilde{y}d = y^* + d(\tilde{y} - y^*)$. Therefore, the latent continuous variable y^* is observed with nonclassical error $u \equiv d(\tilde{y} - y^*)$, which is not independent of y^*. In the context of discrete variables, data contamination is also equivalent to classification error, but contamination also applies to the more general case of continuous variables. Nevertheless, there are still subtle differences between contaminated data models and measurement error models. In contaminated data models, some observations are contaminated, but others are clean. In measurement error models, all observations can be measured with errors of different magnitudes.

Use $Q \equiv Q(y)$ to denote the distribution of y, and use P^* and $\tilde{P}$ to denote the marginal distributions of y^* and $\tilde{y}$. In addition, $P_j^* \equiv P(y^* | d = j)$ and $\tilde{P}_j \equiv P(\tilde{y} | d = j)$ are the distributions of the variables y^* and $\tilde{y}$ conditional on d $(j = 0, 1)$. The marginal probability of data contamination is denoted: $p = P(d = 1)$. Note that in general d is not required to be independent of y^*.

The distribution that can be observed from the data is: $Q = (1-p)P_0^* + p\tilde{P}_1$. But the object of estimation interest is the marginal distribution of the true response: $P^* = (1-p)P_0^* + pP_1^*$. The data does not reveal complete information about P^*. In general, if no prior information about the misclassification probability p is available, the observable distribution Q does not impose any restrictions on the unobservable distribution.

In general, one can assume that the probability of the data error, p, is bounded from above by some constant $\lambda < 1$. In general, one can obtain a consistent estimate of λ and conduct the analysis assuming exact prior information about this parameter.

The following discussion outlines the procedure to obtain a set of probability distributions containing the distribution of the true response for a given λ. Suppose first that for a given p we are given bounds Ψ for P^*, P_1^* and $\tilde{P}_1$. Then the implied distribution of the true response given that we are observing the response without an error is such that: $P_0^* \in \Psi_0^*(p) \equiv \Psi \cap \{(Q - p\tilde{\psi}_1)/(1-p) : \tilde{\psi}_1 \in \Psi\}$. This defines the set of distributions $\Psi_0^*(p)$ where we should expect to find the true response of the error-free observation. Based on this set, for a given probability p, we can confine the marginal distribution of the true response to the following set:

$$
\begin{aligned}
& P^* \in \Psi^*(p) \\
& \equiv \Psi \cap \{(1-p)\psi_0^* + p\psi_1^* : (\psi_0^*, \psi_1^*) \in \Psi_0^*(p) \times \Psi\} \\
& = \Psi \cap \{Q - p\tilde{\psi}_1 + p\psi_1^* : (\tilde{\psi}_1, \psi_1^*) \in \Psi \times \Psi\}.
\end{aligned}
$$

The authors show that the set $\Psi_0^*(p)$ is inside the set $\Psi^*(p)$.

The sets defined above are monotonically increasing in the argument p: for $\delta > 0$: $\Psi^*(p) \subset \Psi^*(p+\delta)$ and $\Psi_0^*(p) \subset \Psi_0^*(p+\delta)$. The fact that $p \leq \lambda < 1$ therefore implies that $P_0^* \in \Psi_0^*(\lambda)$ and $P^* \in \Psi^*(\lambda)$.

If we are willing to assume that d is independent of y^*, so that $P^* = P_1^*$, then the set $\Psi^*(\lambda)$ which contains P^* coincides with the smaller set $\Psi_0^*(\lambda)$.

The identification results for the marginal distribution of the binary outcome are used to place bounds on a general parameter $\tau(\cdot)$, considered as a real-valued function defined on the family of distributions Ψ. Denote the set of values of the general parameter defined on the set of probability distributions Ψ by $T = (T_L, T_U)$. The identified set of parameter values as a function of the set of distributions of the true response y^* can be written as: $\tau(P_0^*) \in T_0^*(\lambda) \equiv \{\tau(\psi): \psi \in \Psi_0^*(\lambda)\}$ and $\tau(P^*) \in T^*(\lambda) \equiv \{\tau(\psi): \psi \in \Psi^*(\lambda)\}$. The shape of these sets depends on λ. Let $T_{0L}^*(\lambda)$ and $T_{0U}^*(\lambda)$ denote the lower and upper bound of $T_0^*(\lambda)$. Let $T_L^*(\lambda)$ and $T_U^*(\lambda)$ denote the lower and upper bounds of $T^*(\lambda)$. The maximum probabilities λ of observing the erroneous outcome such that the estimator $\tau(\cdot)$ is not shifted to the boundary of its range are respectively,

$$\lambda_0^* \equiv \sup\{\lambda : T_L < T_{0L}^*(\lambda) \leq T_{0U}^*(\lambda) < T_U\},$$

and

$$\lambda^* \equiv \sup\{\lambda : T_L < T_L^*(\lambda) \leq T_U^*(\lambda) < T_U\}.$$

The authors call these values λ^* and λ_0^* the "identification breakdown" points of $\tau(\cdot)$, a notation similar to that used in the traditional robust estimation literature in statistics. At these values of the probability of data corruption, the information from the data does not allow us to improve on the precision of a priori information.

A special case of $\tau(\cdot)$ is when $\tau(P_0^*) = P_0^*(A)$ and $\tau(P^*) = P^*(A)$ for a given set A. The general results imply that $P_0^*(A) \in \Psi_0^*(A;\ \lambda) \equiv [0,1] \cap \left[\frac{Q(A)-\lambda}{1-\lambda}, \frac{Q(A)}{1-\lambda}\right]$ and $P^*(A) \in \Psi^*(A;\lambda) \equiv [0,1] \cap [Q(A)-\lambda, Q(A)+\lambda]$. As long as λ is small enough, the identified set of $Q(A)$ is more informative than the entire interval of $[0,1]$.

Another application of the general parameter is bounding the α-quantiles of P_0^* and $P^*: q_0^*(\alpha) = \inf\{t : P_0^*\ [-\infty,\ t] \geq \alpha\}$ and $q^*(\alpha) = \inf\{t : P^*\ [-\infty,\ t] \geq \alpha\}$. Define $r(\gamma)$ to be the γth quantile of Q when $0 \leq \gamma \leq 1$, with the convention that $r(\gamma) = -\infty$ if $\gamma \leq 0$ and $r(\gamma) = \infty$ if $\gamma \leq 1$. Then $q_0^*(\alpha) \in [r(\alpha(1-\lambda)), r(\alpha(1-\lambda)+\lambda)]$ and $q^*(\alpha) \in [r(\alpha-\lambda), r(\alpha+\lambda)]$. These bounds become wider when the probability λ increases.

A parameter $\tau(\cdot)$ is said to be stochastically increasing if $\tau(F) \geq \tau(G)$ when the distribution F first-order stochastically dominates the distribution G. Using the above definition of the quantile function $r(\cdot)$, define $L_\lambda[-\infty, t]$ as $Q[-\infty, t]/(1-\lambda)$ if $t < r(1-\lambda)$, and as 1 if $t \geq r(1-\lambda)$. Also define $U_\lambda[-\infty, t]$ as 0 if $t < r(\lambda)$, and as $(Q[-\infty, t]-\lambda)/(1-\lambda)$ if $t \geq r(\lambda)$. The bounds for the parameter $\tau(\cdot)$ as a function of P_0^* can be stated as $\tau(P_0^*) \in [\tau(L_\lambda), \tau(U_\lambda)]$.

Some stochastic order consistent parameters take the form of $\tau(\psi) = \int g(y)\, d\psi$ for some distribution ψ of the outcome and a function $g(\cdot)$ with a limit at positive infinity equal to K_1 and a limit at negative infinity equal to K_0. In this case, the bounds for the general parameters are respectively $\tau(P_0^*) \in \left[\int g(y)dL_\lambda,\ \int g(y)dU_\lambda\right]$ and $\tau(P^*) \in \left[(1-\lambda)\int g(y)dL_\lambda + \lambda K_0, (1-\lambda)\int g(y)dU_\lambda + \lambda K_1\right]$. Because the bounds for the estimator on the set of distributions of the true outcome given that the outcome is observed without an error do not depend on the asymptotic values of the function $g(\cdot)$, sharp bounds for the estimator of $\tau(P_0^*)$ can be obtained even if the kernel function $g(\cdot)$ is unbounded.

Horowitz and Manski (1995) also show how to provide local bounds for the estimator in the case of smooth functionals of the underlying distributions, and apply their methodology for evaluation of the bounds of the identified set of distributions to analyze the income distribution in the United States. The authors use the data from the Current Population Survey and analyze the characteristics of the income distribution. Approximately 8 percent of the survey respondents provided incomplete data about their income and 4.5 percent of the respondents in the Current Population Survey were not interviewed. This allows the authors to provide a consistent estimate of the upper bound on the probability of erroneous response of 12.1 percent. The application of their method to this data set then allows them to obtain the bounds for the error-corrected quantiles of the income distribution.

Implementing statistical inference for an estimator of bounds or identified set of parameters, such as the estimator of Horowitz and Manski (1995), is an open question that is the focus of active current research. Many recent papers, including Victor Chernozhukov, Hong, and Tamer (2007), investigate the statistical properties of set estimators.

The bounding method of Horowitz and Manski (1995) focuses entirely on estimating the marginal distribution of a random variable that is subject to data contamination and corruption. V. Joseph Hotz, Charles H. Mullin, and Seth G. Sanders (1997) adapt this approach to estimate the causal effect of a dichotomous treatment, which combines information from two marginal distributions of the treated outcome and untreated outcome. They are interested in the effect of teen childbearing on the education and labor market attainment of women, which is difficult because of the selection bias that teen women more likely to bear children might have adverse habits that also negatively affect career attainments. If miscarriage is truly random, it can be used as an instrument to form a valid control group in a natural experiment to estimate the effect of teen childbearing. However, miscarriage has been related by medical evidence to smoking and drug use and other adverse health behavior, and is not necessarily random. Hotz, Mullin, and Sanders treat the observed miscarriage information as contaminated data for a latent random miscarriage indicator: the observed miscarriage is random with certain probability and is nonrandom with the remaining probability. They formulate bounds for the effect of childbearing under this contaminated instrument assumption and obtained very important substantial results. For example, they found that while the ordinary least squares estimate indicates that teenage childbearing lower a mother's annual earnings by $3,016, this estimator can be rejected at any conventional level of significance even when the exclusion restriction on latent-abortion types is not invoked.

Francesca Molinari (2005) uses a different approach to identifying the true outcome distribution from error-contaminated observations. Her direct misclassification approach connects the true and the observed responses by a system of linear equations with the coefficients equal to the misclassification probabilities. Prior information that can tighten up confidence intervals for various statistics of the true response can be incorporated in the form of functional restrictions on the elements of the matrix of misclassification probabilities.

In her model, the set of possible outcome values Y is discrete and the supports of both $\tilde{y}$ and y^* are in Y. The marginal distributions of the true outcome and the observed outcome can be expressed in vector forms $\mathbf{P}^* = [P_j^*, j \in Y] \equiv [\Pr(y^* = j), j \in Y]$ and $\mathbf{Q} = [Q^j, j \in Y] \equiv [\Pr(y = j), j \in Y]$. In addition, define the matrix of conditional probabilities for the observable response given the true response as $\Pi^\star = (\pi_{ij})_{i,j\in Y} \equiv (\Pr(y = i \mid y^* = j))_{i,j\in Y}$. The parameter of interest is a real-valued function on the space of probability distributions $\Psi : \tau[\mathbf{P}^*]$.

The marginal distribution of the observable outcome is related to the marginal distribution of the true outcome through the matrix relation $\mathbf{Q} = \Pi^\star \cdot \mathbf{P}^*$. If the matrix of probabilities $\Pi^\star$ is known and has full rank, than we can retrieve the parametric $\tau(\mathbf{P}^*)$ from the probabilities of the observed outcome by inverting the matrix $\Pi^\star$. This is usually not the case and the prior information comes in the form of the set of possible values of misclassification probabilities $H[\Pi^\star]$ for each element of this matrix.

The identification region for the distribution of the true outcome can be obtained from the observable distribution and the identification bounds as a set:

$$\Psi^{*} = \{\psi : \mathbf{Q} = \Pi \cdot \psi, \Pi \in H[\Pi^{\star}]\}.$$

Given that ψ is a point in the identified set of distributions Ψ^{*}, the identification region for the statistic $\tau(\mathbf{P}^{*})$ can be expressed as: $T^{*} = \{\tau(\psi) : \psi \in \Psi^{*}\}$. We split the set of identifying constraints into two components. $H^{P}[\Pi^{\star}]$ denotes the set of matrices satisfying the probabilistic constraints and $H^{E}[\Pi^{\star}]$ denotes the set of matrices satisfying the constraints from validation studies. The identified set for the matrix of misclassification probabilities is: $H[\Pi^{\star}] = H^{P}[\Pi^{\star}] \cap H^{E}[\Pi^{\star}]$. Apparently, the geometry of the set $H[\Pi^{\star}]$ will be translated to the geometry of the set Ψ^{*}. In particular, if the set of restrictions from validation studies is not connected, then the identified set of the statistic $\tau(\cdot)$ can also be disconnected.

The set of probabilistic restrictions implies that the matrix $\Pi^{\star}$ should be stochastic, in the sense that each column lies on an unit simplex, and that multiplication of the vector of probabilities $\mathbf{P}^{*}$ by this matrix should give a proper distribution $\mathbf{Q}$. If Δ_n is an n-dimensional simplex and $\mathrm{conv}(a_1, a_2, \ldots, a_n)$ is the convex hull of a collection of vectors $\{a_k\}_{k=1}^{n}$, the set of probabilistic restrictions can be written as

$$\begin{aligned} H^{P}[\Pi^{\star}] = \Big\{\Pi : \pi_j \in \Delta_{|Y|-1} \quad &\text{and} \\ \psi_j^{*} \geq 0, \forall j \in Y, \psi^{*} \in \Psi^{*}, \quad &\text{and} \\ \mathbf{Q} \in \mathrm{conv}(\pi_1, \ldots, \pi_{|Y|})\Big\}&, \end{aligned}$$

where π_j stands for the jth column of matrix $\Pi^{\star}$.

The author provides different examples of a possible set of validation restrictions. One such restriction has been considered in Horowitz and Manski (1995) outlined above, where one can impose an upper bound on the probabilities of erroneous outcome, and thus impose a lower bound restriction on the elements of the diagonal of the matrix $\Pi^{\star}$. The other example of such restrictions is when the variable y_0 tends to be overreported which means that

$$H^{E}[\Pi^{\star}] = \{\Pi : \pi_{ij} = 0, \forall i < j \in Y\}.$$

The set of probabilistic restrictions is usually not convex because of the set of validation restrictions. The resulting identification set for the elements of $\Pi^{\star}$ can be convex or nonconvex, connected or disconnected.

The author then uses the technique for estimation of set-identified parameters to recover the elements of the true outcome distribution $\mathbf{P}^{*}$. The technique is based on treatment of restrictions for the elements in $H^{E}[\Pi^{\star}]$ as inequality and equality constraints given the probabilistic restrictions. Then the problem of verifying whether an element $\psi^{*} \in \Psi^{*}$ satisfies the constraints reduces to the problem of looking for a feasible solution in a linear programming problem. The author proves that the identified region constructed in this way will be consistent in the supremum-norm sense for the true identified region.

Molinari (2005) uses the data from the Health and Retirement Study to illustrate her identification and estimation methodology. The author studies the distribution of the types of pension plans in the population of the currently employed Americans for the period between the years 1992 and 1998. A significant inference problem is that, in general, the workers might be misinformed about the characteristics of their pension plans and, for this reason, a substantial amount of error might be present in the survey data. The respondents have three pension plans available, and the author possesses an additional dataset that matches the individuals in the survey to the exact data provided by the Social Security Administration. This additional dataset is used to impose the restrictions on the matrix of misreporting probabilities. Then, assuming stability of the distribution of misreporting probabilities, the author obtains the confidence sets for the pension plan choice probabilities for individuals in the three survey subsamples for three different periods of time.

6. *Conclusion*

In summary, inferences based on linear models and classical measurement errors can be very specialized and nonrobust. Therefore, in this survey we have focused on the recent advances in identification and estimation of nonlinear EIV models with classical measurement errors and nonlinear EIV models with nonclassical measurement errors, as well as some results on partial identification in nonlinear EIV models. We have briefly discussed the applications of various new methods immediately after the methods are introduced. Additional applications using econometric techniques for solving measurement error problems can be found in Carroll et al. (2006), Bound, Brown, and Mathiowetz (2001), and Ridder and Moffit (2007).

Due to the lack of time and space, we have not reviewed many papers on measurement errors in detail. We have not mentioned any Bayesian approach to measurement error problems. We have not discussed methods to solve measurement errors problems that take advantage of panel data and time series structures; see, e.g., Hsiao (1991), Horowitz and Markatou (1996), Karen E. Dynan (2000), and Jonathan A. Parker and Bruce Preston (2005) for such applications. We have also not discussed the literature on small noise approximation to assess the effect of measurement errors; see, e.g., Chesher (1991), Chesher and Schluter (2002), and Chesher, Dumangane, and Smith (2002).

Despite numerous articles that have been written on the topic of measurement errors in econometrics and statistics over the years, there are still many unsolved important questions. For example, the implications of measurement errors and data contamination on complex (nonlinear) structural models in labor economics, industrial organization, and asset pricing are yet to be understood and studied. Also, it is often the case that not all mismeasured variables are validated in auxiliary data sets; hence how to make use of partial information in validation studies is an important question. Finally, there is relatively little work on the problem of misspecification of various crucial identifying assumptions for nonlinear EIV models.

References

Ai, Chunrong, and Xiaohong Chen. 2003. "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions." *Econometrica*, 71(6): 1795–1843.

Amemiya, Yasuo. 1985. "Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model." *Journal of Econometrics*, 28(3): 273–89.

Bollinger, Christopher R. 1998. "Measurement Error in the Current Population Survey: A Nonparametric Look." *Journal of Labor Economics*, 16(3): 576–94.

Bonhomme, Stephane, and Jean-Marc Robin. 2010. "Generalized Non-parametric Deconvolution with an Application to Earnings Dynamics." *Review of Economic Studies*, 77(2): 491–533.

Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers. 1994. "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, 12(3): 345–68.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics*, *Volume 5*, ed. James J. Heckman and Edward Leamer, 3705–3843. Amsterdam; London and New York: Elsevier Science, North-Holland.

Bound, John, and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9(1): 1–24.

Butucea, Cristina, and Marie-Luce Taupin. 2008. "New *M*-Estimators in Semi-parametric Regression with Errors in Variables." *Annales de l'Institut Henri Poincaré—Probabilités et Statistiques*, 44(3): 393–421.

Carroll, Raymond J., and Peter Hall. 1988. "Optimal Rates of Convergence for Deconvolving a Density." *Journal of the American Statistical Association*, 83(404): 1184–86.

Carroll, Raymond J., David Ruppert, Ciprian M.

Crainiceanu, Tor D. Tosteson, and Margaret R. Karagas. 2004. "Nonlinear and Nonparametric Regression and Instrumental Variables." *Journal of the American Statistical Association*, 99(467): 736–50.

Carroll, Raymond J. David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective,* Second Edition. Boca Raton: Taylor and Francis, Chapman and Hall.

Carroll, Raymond J., and Leonard A. Stefanski. 1990. "Approximate Quasi-likelihood Estimation in Models with Surrogate Predictors." *Journal of the American Statistical Association*, 85(411): 652–63.

Carroll, Raymond J., and M. P. Wand. 1991. "Semiparametric Estimation in Logistic Measurement Error Models." *Journal of the Royal Statistical Society, Series B*, 53(3): 573–85.

Chen, Xiaohong, Han Hong, and Elie Tamer. 2005. "Measurement Error Models with Auxiliary Data." *Review of Economic Studies*, 72(2): 343–66.

Chen, Xiaohong, Han Hong, and Alessandro Tarozzi. 2008. "Semiparametric Efficiency in GMM Models with Auxiliary Data." *Annals of Statistics*, 36(2): 808–43.

Chen, Xiaohong, and Yingyao Hu. 2006. "Identification and Inference of Nonlinear Models Using Two Samples with Arbitrary Measurement Errors." Cowles Foundation Discussion Paper 1590.

Chernozhukov, Victor, Han Hong, and Elie Tamer. 2007. "Estimation and Confidence Regions for Parameter Sets in Econometric Models." *Econometrica*, 75(5): 1243–84.

Chesher, Andrew. 1991. "The Effect of Measurement Error." *Biometrika*, 78(3): 451–62.

Chesher, Andrew, Montezuma Dumangane, and Richard J. Smith. 2002. "Duration Response Measurement Error." *Journal of Econometrics*, 111(2): 169–94.

Chesher, Andrew, and Christian Schluter. 2002. "Welfare Measurement and Measurement Error." *Review of Economic Studies*, 69(2): 357–78.

Devereux, Paul J., and Gautam Tripathi. 2009. "Optimally Combining Censored and Uncensored Datasets." *Journal of Econometrics*, 151(1): 17–32.

Dickens, William T., and Brian A. Ross. 1984. "Consistent Estimation Using Data from More Than One Sample." National Bureau of Economic Research Technical Working Paper 33.

Dynan, Karen E. 2000. "Habit Formation in Consumer Preferences: Evidence from Panel Data." *American Economic Review*, 90(3): 391–406.

Fan, Jianqing. 1991. "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems." *Annals of Statistics*, 19(3): 1257–72.

Fan, Jianqing, and Young K. Truong. 1993. "Nonparametric Regression with Errors in Variables." *Annals of Statistics*, 21(4): 1900–1925.

Fan, Yanqin, and Qi Li. 1996. "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms." *Econometrica*, 64(4): 865–90.

Friedman, Milton. 1957. *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Frisch, Ragnar. 1934. *Statistical Conference Study*. Oslo: University Institute of Economics.

Fuller, Wayne A. 1987. *Measurement Error Models*. New York: Wiley.

Gini, Corrado. 1921. "Sull' interpolazione di una retta quando i valori variabile indipendente sono affetti da errori accidentali." *Metron*, 1(3): 63–82.

Griliches, Zvi, and Vidar Ringstad. 1970. "Error-in-the-Variables Bias in Nonlinear Contexts." *Econometrica*, 38(2): 368–70.

Hahn, Jinyong. 1998. "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica*, 66(2): 315–31.

Han, Aaron K. 1987. "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator." *Journal of Econometrics*, 35(2–3): 303–16.

Hausman, Jerry A. 2001. "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives*, 15(4): 57–67.

Hausman, Jerry A., Jason Abrevaya, and F. M. Scott-Morton. 1998. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics*, 87(2): 239–69.

Hausman, Jerry A., Whitney K. Newey, Hidehiko Ichimura, and James L. Powell. 1991. "Identification and Estimation of Polynomial Errors-in-Variables Models." *Journal of Econometrics*, 50(3): 273–95.

Hausman, Jerry A., Whitney K. Newey, and James L. Powell. 1995. "Nonlinar Errors in Variables Estimation of Some Engel Curves." *Journal of Econometrics*, 65(1): 205–33.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 71(4): 1161–89.

Hong, Han, and Elie Tamer. 2003. "A Simple Estimator for Nonlinear Error in Variable Models." *Journal of Econometrics*, 117(1): 1–19.

Horowitz, Joel L., and Wolfgang Hardle. 1994. "Testing a Parametric Model against a Semiparametric Alternative." *Econometric Theory*, 10(5): 821–48.

Horowitz, Joel L., and Charles F. Manski. 1995. "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica*, 63(2): 281–302.

Horowitz, Joel L., and Marianthi Markatou. 1996. "Semiparametric Estimation of Regression Models for Panel Data." *Review of Economic Studies*, 63(1): 145–68.

Hotz, V. Joseph, Charles H. Mullin, and Seth G. Sanders. 1997. "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analysing the Effects of Teenage Childbearing." *Review of Economic Studies*, 64(4): 575–603.

Hsiao, Cheng. 1991. "Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data." *Review of Economic Studies*, 58(4): 717–31.

Hu, Yingyao. 2008. "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution." *Journal of Econometrics*, 144(1): 27–61.

Hu, Yingyao, and Susanne Schennach. 2008. "Instrumental Variable Treatment of Nonclassical Error Models." *Econometrica*, 76(1): 195–216.

Hu, Yingyao, and Geert Ridder. Forthcoming. "Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information." *Journal of Applied Econometrics*.

Hwang, Gene T., and Leonard A. Stefanski. 1994. "Monotonicity of Regression Functions in Structural Measurement Error Models." *Statistics & Probability Letters*, 20(2): 113–16.

Ichimura, Hidehiko, and Elena Martinez-Sanchis. 2005. "Identification and Estimation of GMM Models by Combining Two Data Sets." Unpublished.

Kane, Thomas J., Cecelia Elena Rouse, and Douglas Staiger. 1999. "Estimating Returns to Schooling When Schooling Is Misreported." National Bureau of Economic Research Working Paper 7235.

Klepper, Steven, and Edward E. Leamer. 1984. "Consistent Sets of Estimates for Regressions with Errors in All Variables." *Econometrica*, 52(1): 163–83.

Lavergne, Pascal, and Quang Vuong. 2000. "Nonparametric Significance Testing." *Econometric Theory*, 16(4): 576–601.

Lee, Lung-fei, and Jungsywan H. Sepanski. 1995. "Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data." *Journal of the American Statistical Association*, 90(429): 130–40.

Lewbel, Arthur. 1997. "Constructing Instruments for Regressions with Measurement Error When No Additional Data Are Available, with an Application to Patents and R&D." *Econometrica*, 65(5): 1201–13.

Lewbel, Arthur. 2007. "Estimation of Average Treatment Effects with Misclassification." *Econometrica*, 75(2): 537–51.

Li, Tong. 2002. "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models." *Journal of Econometrics*, 110(1): 1–26.

Li, Tong, and Cheng Hsiao. 2004. "Robust Estimation of Generalized Linear Models with Measurement Errors." *Journal of Econometrics*, 118(1–2): 51–65.

Li, Tong, Isabelle Perrigne, and Quang Vuong. 2000. "Conditionally Independent Private Information in OCS Wildcat Auctions." *Journal of Econometrics*, 98(1): 129–61.

Li, Tong, and Quang Vuong. 1998. "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators." *Journal of Multivariate Analysis*, 65(2): 139–65.

Linton, Oliver, and Yoon-Jae Whang. 2002. "Nonparametric Estimation with Aggregated Data." *Econometric Theory*, 18(2): 420–68.

Mahajan, Aprajit. 2006. "Identification and Estimation of Regression Models with Misclassification." *Econometrica*, 74(3): 631–65.

Molinari, Francesca. 2005. "Partial Identification of Probability Distributions with Misclassified Data." Cornell University Center for Analytic Economics Working Paper 05-10.

Newey, Whitney K. 2001. "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models." *Review of Economics and Statistics*, 83(4): 616–27.

Parker, Jonathan A., and Bruce Preston. 2005. "Precautionary Saving and Consumption Fluctuations." *American Economic Review*, 95(4): 1119–43.

Rao, P. 1992. *Identifiability in Stochastic Models: Characterization of Probability Distributions*. Boston and London: Academic Press.

Reiersøl, Olav. 1950. "Identifiability of a Linear Relation between Variables Which Are Subject to Error." *Econometrica*, 18(4): 375–89.

Ridder, Geert, and Robert Moffitt. 2007. "The Econometrics of Data Combination." In *Handbook of Econometrics, Volume 6B*, ed. James J. Heckman and Edward Leamer, 5469–5548. Amsterdam and Oxford: Elsevier, North-Holland.

Robins, James M., Steven D. Mark, and Whitney K. Newey. 1992. "Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders." *Biometrics*, 48(2): 479–95.

Schennach, Susanne M. 2004a. "Estimation of Nonlinear Models with Measurement Error." *Econometrica*, 72(1): 33–75.

Schennach, Susanne M. 2004b. "Nonparametric Regression in the Presence of Measurement Error." *Econometric Theory*, 20(6): 1046–93.

Schennach, Susanne M. 2007. "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models." *Econometrica*, 75(1): 201–39.

Sepanski, Jungsywan H., and Raymond J. Carroll. 1993. "Semiparametric Quasilikelihood and Variance Function Estimation in Measurement Error Models." *Journal of Econometrics*, 58(1–2): 223–56.

Sherman, Robert P. 1993. "The Limiting Distribution of the Maximum Rank Correlation Estimator." *Econometrica*, 61(1): 123–37.

Taupin, Marie-Luce. 2001. "Semi-parametric Estimation in the Nonlinear Structural Errors-in-Variables Model." *Annals of Statistics*, 29(1): 66–93.

Wang, Liqun. 2004. "Estimation of Nonlinear Models with Berkson Measurement Errors." *Annals of Statistics*, 32(6): 2559–79.

Wang, Liqun, and Cheng Hsiao. 1995. "Simulation-Based Semiparametric Estimation of Nonlinear Errors-in-Variables Models." Unpublished.

Wansbeek, Tom, and Erik Meijer. 2000. *Measurement Error and Latent Variables in Econometrics*. Amsterdam; London and New York: Elsevier Science, North-Holland.

Zinde-Walsh, Victoria. 2007. "Errors-in-Variables Models: A Generalized Functions Approach." Unpublished.

This article has been cited by:

1. Hao Dong, Taisuke Otsu, Luke Taylor. 2020. AVERAGE DERIVATIVE ESTIMATION UNDER MEASUREMENT ERROR. *Econometric Theory* **58**, 1-30. [Crossref]
2. Julio Acuna. 2020. Non-classical measurement error with false positives and negatives. *Applied Economics Letters* **2005**, 1-6. [Crossref]
3. Tanya Byker, Italo Gutierrez. 2020. Estimating the Impact of Sterilization Under a Government Campaign Using Contaminated Treatment Data. *Economic Development and Cultural Change* . [Crossref]
4. Ruth H. Keogh, Pamela A. Shaw, Paul Gustafson, Raymond J. Carroll, Veronika Deffner, Kevin W. Dodd, Helmut Küchenhoff, Janet A. Tooze, Michael P. Wallace, Victor Kipnis, Laurence S. Freedman. 2020. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—Basic theory and simple methods of adjustment. *Statistics in Medicine* **39**:16, 2197-2231. [Crossref]
5. Ling Leng, Wei Zhu. 2020. Compound Regression and Constrained Regression: Nonparametric Regression Frameworks for EIV Models. *The American Statistician* **74**:3, 226-232. [Crossref]
6. Dan Ben-Moshe. 2020. IDENTIFICATION OF LINEAR REGRESSIONS WITH ERRORS IN ALL VARIABLES. *Econometric Theory* **62**, 1-31. [Crossref]
7. Yaman Hamed, Ahmed Ibrahim Alzahrani, A'fza Shafie, Zahiraniza Mustaffa, Mokhtar Che Ismail, Kee Kok Eng. 2020. Two steps hybrid calibration algorithm of support vector regression and K-nearest neighbors. *Alexandria Engineering Journal* **59**:3, 1181-1190. [Crossref]
8. Bingyang Lv, Yongzheng Liu, Yan Li. 2020. Fiscal incentives, competition, and investment in China. *China Economic Review* **59**, 101371. [Crossref]
9. Andrew Chesher, Adam M. Rosen. Generalized instrumental variable models, methods, and applications . [Crossref]
10. Christos Makridis, Cary Wu. 2020. Ties That Bind (and Social Distance): How Social Capital Helps Communities Weather the COVID-19 Pandemic. *SSRN Electronic Journal* . [Crossref]
11. Yao Luo, Ruli Xiao. 2020. Identification of Auction Models Using Order Statistics. *SSRN Electronic Journal* . [Crossref]
12. Susanne M. Schennach. Mismeasured and unobserved variables . [Crossref]
13. Christos Makridis, Cary Wu. 2020. Ties that Bind (and Social Distance): How Social Capital Helps Communities Weather the COVID-19 Pandemic. *SSRN Electronic Journal* . [Crossref]
14. Karim Chalak. 2020. Nonparametric Gini-Frisch Bounds. *SSRN Electronic Journal* . [Crossref]
15. Julio Acuna. 2020. Non-classical Measurement Error with False Positives and Negatives. *SSRN Electronic Journal* . [Crossref]
16. Christos Makridis, Tao Wang. 2020. Learning from Friends in a Pandemic: Social Networks and the Macroeconomic Response of Consumption. *SSRN Electronic Journal* . [Crossref]
17. Kengo Kato, Yuya Sasaki. 2019. Uniform confidence bands for nonparametric errors-in-variables regression. *Journal of Econometrics* **213**:2, 516-555. [Crossref]
18. Alexandre Poirier, Nicolas L. Ziebarth. 2019. Estimation of Models With Multiple-Valued Explanatory Variables. *Journal of Business & Economic Statistics* **37**:4, 586-597. [Crossref]

19. Takahide Yanagi. 2019. Inference on local average treatment effects for misclassified treatment. *Econometric Reviews* **38**:8, 938-960. [Crossref]
20. Isabelle Perrigne, Quang Vuong. 2019. Econometrics of Auctions and Nonlinear Pricing. *Annual Review of Economics* **11**:1, 27-54. [Crossref]
21. Karim Chalak, Daniel Kim. 2019. Measurement Error Without the Proxy Exclusion Restriction. *Journal of Business & Economic Statistics* **5**, 1-17. [Crossref]
22. Benjamin Williams. 2019. Identification of a nonseparable model under endogeneity using binary proxies for unobserved heterogeneity. *Quantitative Economics* **10**:2, 527-563. [Crossref]
23. Gordon Anderson. Comparing Latent Subgroups 135-151. [Crossref]
24. Karun Adusumilli, Taisuke Otsu. 2018. NONPARAMETRIC INSTRUMENTAL REGRESSION WITH ERRORS IN VARIABLES. *Econometric Theory* **34**:6, 1256-1280. [Crossref]
25. Kengo Kato, Yuya Sasaki. 2018. Uniform confidence bands in deconvolution with unknown error distribution. *Journal of Econometrics* **207**:1, 129-161. [Crossref]
26. Dennis Glennon, Hua Kiefer, Tom Mayock. 2018. Measurement error in residential property valuation: An application of forecast combination. *Journal of Housing Economics* **41**, 1-29. [Crossref]
27. Arun Advani, Bansi Malde. 2018. CREDIBLY IDENTIFYING SOCIAL EFFECTS: ACCOUNTING FOR NETWORK FORMATION AND MEASUREMENT ERROR. *Journal of Economic Surveys* **32**:4, 1016-1044. [Crossref]
28. Martin Karlsson, Nicolas R. Ziebarth. 2018. Population health effects and health-related costs of extreme temperatures: Comprehensive evidence from Germany. *Journal of Environmental Economics and Management* **91**, 93-117. [Crossref]
29. Robert Grafstein. 2018. The problem of polarization. *Public Choice* **176**:1-2, 315-340. [Crossref]
30. Mitsukuni Nishida, Marc Remer. 2018. The Determinants and Consequences of Search Cost Heterogeneity: Evidence from Local Gasoline Markets. *Journal of Marketing Research* **55**:3, 305-320. [Crossref]
31. Hachmi Ben Ameur, Fredj Jawadi, Abdoulkarim Idi Cheffou, Wael Louhichi. 2018. Measurement errors in stock markets. *Annals of Operations Research* **262**:2, 287-306. [Crossref]
32. Aaron Chalfin, Justin McCrary. 2018. Are U.S. Cities Underpoliced? Theory and Evidence. *The Review of Economics and Statistics* **100**:1, 167-186. [Crossref]
33. Christoph Breunig, Enno Mammen, Anna Simoni. 2018. Nonparametric estimation in case of endogenous selection. *Journal of Econometrics* **202**:2, 268-285. [Crossref]
34. Irene Botosaru, Federico H. Gutierrez. 2018. Difference-in-differences when the treatment status is observed in only one period. *Journal of Applied Econometrics* **33**:1, 73-90. [Crossref]
35. Yingyao Hu, Zhongjian Lin. 2018. Misclassification and the Hidden Silent Rivalry. *SSRN Electronic Journal* . [Crossref]
36. Yulai Zhang, Weifeng Ma, Guiming Luo. 2017. Inferring causal directions from uncertain data. *Engineering Applications of Artificial Intelligence* **65**, 261-267. [Crossref]
37. Yingyao Hu, Tom Wansbeek. 2017. Measurement error models: Editors' introduction. *Journal of Econometrics* **200**:2, 151-153. [Crossref]

38. Yingyao Hu. 2017. The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics. *Journal of Econometrics* **200**:2, 154-168. [Crossref]

39. Jinyong Hahn, Geert Ridder. 2017. Instrumental variable estimation of nonlinear models with nonclassical measurement error using control variables. *Journal of Econometrics* **200**:2, 238-250. [Crossref]

40. Laurent Davezies, Thomas Le Barbanchon. 2017. Regression discontinuity design with continuous measurement error in the running variable. *Journal of Econometrics* **200**:2, 260-281. [Crossref]

41. Bruce D. Meyer, Nikolas Mittag. 2017. Misclassification in binary choice models. *Journal of Econometrics* **200**:2, 295-311. [Crossref]

42. Xiaohong Chen, Oliver Linton, Yanping Yi. 2017. Semiparametric identification of the bid–ask spread in extended Roll models. *Journal of Econometrics* **200**:2, 312-325. [Crossref]

43. Erich Battistin, Michele De Nadai, Daniela Vuri. 2017. Counting rotten apples: Student achievement and score manipulation in Italian elementary Schools. *Journal of Econometrics* **200**:2, 344-362. [Crossref]

44. Yibin Liu, Wenbin Wu. 2017. Closed-form estimation of a regression model with a mismeasured binary regressor and heteroskedasticity. *Statistics & Probability Letters* **125**, 202-206. [Crossref]

45. Giovanni D'Alessio. 2017. Measurement Errors in Consumption Surveys and the Estimation of Poverty and Inequality Indices. *SSRN Electronic Journal* . [Crossref]

46. Takahide Yanagi. 2017. Inference on Local Average Treatment Effects for Misclassified Treatment. *SSRN Electronic Journal* . [Crossref]

47. Takahide Yanagi. 2017. Regression Discontinuity Designs with Nonclassical Measurement Errors. *SSRN Electronic Journal* . [Crossref]

48. Susanne M. Schennach. 2016. Recent Advances in the Measurement Error Literature. *Annual Review of Economics* **8**:1, 341-377. [Crossref]

49. Giovanni Compiani, Yuichi Kitamura. 2016. Using Mixtures in Econometric Models: A Brief Review and Some New Results. *The Econometrics Journal* **19**:3, C95-C127. [Crossref]

50. Marco Batarce. 2016. Estimation of urban bus transit marginal cost without cost data. *Transportation Research Part B: Methodological* **90**, 241-262. [Crossref]

51. Lorenzo Almada, Ian McCarthy, Rusty Tchernis. 2016. What Can We Learn about the Effects of Food Stamps on Obesity in the Presence of Misreporting?. *American Journal of Agricultural Economics* **98**:4, 997-1017. [Crossref]

52. Maxim Pinkovskiy, Xavier Sala-i-Martin. 2016. Lights, Camera … Income! Illuminating the National Accounts-Household Surveys Debate *. *The Quarterly Journal of Economics* **131**:2, 579-631. [Crossref]

53. Michele De Nadai, Arthur Lewbel. 2016. Nonparametric errors in variables models with measurement errors on both sides of the equation. *Journal of Econometrics* **191**:1, 19-32. [Crossref]

54. Chunmian Ge, Ke-Wei Huang, Ivan P. L. Png. 2016. Engineer/scientist careers: Patents, online profiles, and misclassification bias. *Strategic Management Journal* **37**:1, 232-253. [Crossref]

55. Ji-Liang Shiu. 2016. Identification and estimation of endogenous selection models in the presence of misclassification errors. *Economic Modelling* **52**, 507-518. [Crossref]

56. M. Hakan Satman, Erkin Diyarbakirlioglu. 2015. Reducing errors-in-variables bias in linear regression using compact genetic algorithms. *Journal of Statistical Computation and Simulation* **85**:16, 3216-3235. [Crossref]

57. Alfonso Irarrazabal, Andreas Moxnes, Luca David Opromolla. 2015. The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs. *Review of Economics and Statistics* **97**:4, 777-792. [Crossref]

58. Yingyao Hu, Yuya Sasaki. 2015. Closed-form estimation of nonparametric models with non-classical measurement errors. *Journal of Econometrics* **185**:2, 392-408. [Crossref]

59. Suyong Song. 2015. Semiparametric estimation of models with conditional moment restrictions in the presence of nonclassical measurement errors. *Journal of Econometrics* **185**:1, 95-109. [Crossref]

60. Changha Hwang. 2015. Partially linear support vector orthogonal quantile regression with measurement errors. *Journal of the Korean Data and Information Science Society* **26**:1, 209-216. [Crossref]

61. Yingyao Hu. 2015. Microeconomic Models with Latent Variables: Applications of Measurement Error Models in Empirical Industrial Organization and Labor Economics. *SSRN Electronic Journal* . [Crossref]

62. Alexandre Poirier, Nicolas Ziebarth. 2015. A Simple Estimator for Datasets with Non-Unique Identifiers. *SSRN Electronic Journal* . [Crossref]

63. Dennis Glennon, Hua Kiefer, Tom Mayock. 2015. Housing Value Estimation: An Application of Forecast Combination to Residential Property Valuation. *SSRN Electronic Journal* . [Crossref]

64. Mitsukuni Nishida, Marc Remer. 2015. The Determinants and Consequences of Search Cost Heterogeneity: Evidence from Local Gasoline Markets. *SSRN Electronic Journal* . [Crossref]

65. Francis DiTraglia, Camilo Garcia-Jimeno. 2015. On Mis-Measured Binary Regressors: New Results and Some Comments on the Literature. *SSRN Electronic Journal* . [Crossref]

66. Xiaoli L. Etienne, Scott H. Irwin, Philip Garcia. 2015. Price Explosiveness, Speculation, and Grain Futures Prices. *American Journal of Agricultural Economics* **97**:1, 65-87. [Crossref]

67. Florence Neymotin. 2014. How Parental Involvement Affects Childhood Behavioral Outcomes. *Journal of Family and Economic Issues* **35**:4, 433-451. [Crossref]

68. Victoria Zinde-Walsh. 2014. MEASUREMENT ERROR AND DECONVOLUTION IN SPACES OF GENERALIZED FUNCTIONS. *Econometric Theory* **30**:6, 1207-1246. [Crossref]

69. Seonjin Kim, Zhibiao Zhao. 2014. Specification test for Markov models with measurement errors. *Journal of Multivariate Analysis* **130**, 118-133. [Crossref]

70. Erich Battistin, Michele De Nadai, Barbara Sianesi. 2014. Misreported schooling, multiple measures and returns to educational qualifications. *Journal of Econometrics* **181**:2, 136-150. [Crossref]

71. Joel L. Horowitz. 2014. Ill-Posed Inverse Problems in Economics. *Annual Review of Economics* **6**:1, 21-51. [Crossref]

72. Marc Henry, Yuichi Kitamura, Bernard Salanié. 2014. Partial identification of finite mixtures in econometric models. *Quantitative Economics* **5**:1, 123-144. [Crossref]

73. Erich Battistin, Andrew Chesher. 2014. Treatment effect estimation with covariate measurement error. *Journal of Econometrics* **178**:2, 707-715. [Crossref]

74. Chunmian Ge, Ke-Wei Huang, Ivan P. L. Png. 2014. Engineer/Scientist Careers: Patents, Online Profiles, and Misclassification Bias. *SSRN Electronic Journal* . [Crossref]

75. Chunmian Ge, Ke-Wei Huang, Ivan P. L. Png. 2014. Engineer/Scientist Careers: Patents, Online Profiles, and Misclassification Bias. *SSRN Electronic Journal* . [Crossref]

76. Nicolas R. Ziebarth. 2013. Long-term absenteeism and moral hazard—Evidence from a natural experiment. *Labour Economics* **24**, 277-292. [Crossref]

77. Ryo Kambayashi, Daiji Kawaguchi, Ken Yamada. 2013. Minimum wage in a deflationary economy: The Japanese experience, 1994–2003. *Labour Economics* **24**, 264-276. [Crossref]

78. Alex Eapen. 2013. FDI spillover effects in incomplete datasets. *Journal of International Business Studies* **44**:7, 719-744. [Crossref]

79. Louis-Philippe Morin. 2013. Estimating the benefit of high school for university-bound students: evidence of subject-specific human capital accumulation. *Canadian Journal of Economics/Revue canadienne d'économique* **46**:2, 441-468. [Crossref]

80. Jeffrey F. Timmons, Daniel Broid. 2013. The Political Economy of Municipal Transfers: Evidence from Mexico. *Publius: The Journal of Federalism* **43**:4, 551-579. [Crossref]

81. Jeffrey F. Timmons, Daniel S. Broid. 2013. The Political Economy of Municipal Transfers: Evidence from Mexico. *SSRN Electronic Journal* . [Crossref]

82. Yonghong An, Xun Tang. 2013. Identifying Structural Models of Committee Decisions with Heterogeneous Tastes and Ideological Bias. *SSRN Electronic Journal* . [Crossref]

83. Per Hjertstrand. 2013. A Simple Method to Account for Measurement Errors in Revealed Preference Tests. *SSRN Electronic Journal* . [Crossref]

84. Ida Mariati Hutabarat, Asep Saefuddin, Anik Djuraidah, I Wayan Mangku. 2013. Estimating the Parameters Geographically Weighted Regression (GWR) with Measurement Error. *Open Journal of Statistics* **03**:06, 417-421. [Crossref]

85. Marc Henry, Yuichi Kitamura, Bernard Salanie. 2010. Identifying Finite Mixtures in Econometric Models. *SSRN Electronic Journal* . [Crossref]