Recent Advances in the Measurement Error Literature

Susanne M. Schennach

Department of Economics, Brown University, Providence, Rhode Island 02912;
email: smschenn@alum.mit.edu

Annu. Rev. Econ. 2016. 8:341–77

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi:
10.1146/annurev-economics-080315-015058

JEL codes: C13, C14, C39

Keywords

latent variable, factor, nonlinear, nonclassical, nonseparable

Abstract

This article reviews recent significant progress made in developing estimation and inference methods for nonlinear models in the presence of mismeasured data that may or may not conform to the classical assumption of independent zero-mean errors. The aim is to cover a broad range of methods having differing levels of complexity and strength of the required assumptions. Simple approaches that form the elementary building blocks of more advanced approaches are discussed first. Then, special attention is devoted to methods that rely on readily available auxiliary variables (e.g., repeated measurements, indicators, or instrumental variables). Results relaxing most of the commonly invoked simplifying assumptions are presented (linear measurement structure, independent errors, zero-mean errors, availability of auxiliary information). This article also provides an overview of important connections with related fields, such as latent variable models, nonlinear panel data, factor models, and set identification, and applications of the methods to other fields traditionally unrelated to measurement error models.

1. INTRODUCTION

The statistical analysis of error-contaminated data has a long history, dating back at least to the early days of econometrics (e.g., Frisch 1934). Perhaps surprisingly, however, this topic has remained fairly active, with a recent surge in activity, as researchers seek to relax most of the commonly held assumptions (e.g., linearity, independence). This article emphasizes the recent econometric literature on the topic and mostly centers on the question of the identification and estimation of general nonlinear models with measurement error without simply assuming that the distribution of the measurement error is known or directly observable.

This article is organized as follows. First, the origins of measurement error bias are explained, before describing simple approaches that rely on distributional knowledge regarding the measurement error (e.g., deconvolution or validation data techniques). Methods are then described that secure identification via more readily available auxiliary variables (e.g., repeated measurements, multiple indicators, measurement systems with a factor model structure, instrumental variables, and panel data). An overview of methods exploiting higher-order moments or bounding techniques to avoid the need for auxiliary information is presented next. Special attention is devoted to a recently introduced general method to handle a broad class of latent variable models, called entropic latent variable integration via simulation (ELVIS). Finally, the complex but very active topic of nonclassical measurement error is covered, and applications of measurement error techniques to other fields are outlined.

1.1. Other Reviews

There are other available reviews on the topic of measurement error that deliver a somewhat different focus. In the statistics literature, methods primarily aimed at linear models are discussed by Fuller (1987) and Cheng & Ness (1999). The popular book by Carroll et al. (2006) covers nonlinear measurement error models, with a special focus on bias reduction (also called approximate consistency). Reviews that center on the econometrics literature are also available. Wansbeek & Meijer (2000) also focus primarily on linear models and make direct connections with latent variables and factor models. A broad review of nonlinear measurement error models with an emphasis on the use of auxiliary samples containing error-free covariates is provided by Chen et al. (2011). The present review updates the earlier review of Schennach (2013c), especially regarding latent variables, factor models, and nonseparable error and providing more insight into the connection among different approaches.

1.2. Models and Conventions

Let us first summarize the models and conventions used throughout the article. A lowercase letter denotes a specific value of the random variable (or random vector) denoted by the corresponding uppercase letter. Let $f_V(v)$ and $F_V(v)$ denote, respectively, the density[1] and the distribution of the random variable V and similarly for random vectors, whose dimension is denoted d_V.

The unobserved but true values of the variable of interest X^* and its observed but mismeasured counterpart X are related through

$$X = X^* + \Delta X, \tag{1}$$

[1]This article is mostly phrased in terms of densities (with respect to the Lebesgue measure for continuous variables or the counting measure for discrete ones) for simplicity of exposition, but most identification results hold for general probability measures, at the expense of notational complications.

where ΔX denotes the measurement error. The variables of interest may also include another vector of perfectly measured variables Y (or at least variables for which the measurement error would not induce bias, such as the dependent variable in a standard regression). Very often, identification of the model will rely on the availability of a vector of other observed variables (e.g., repeated measurements or instruments), denoted by Z. When the model includes disturbances in the process generating Y or Z, these are denoted by ΔY or ΔZ, respectively. Any of these variables could be vectors.

For the most part, regularity conditions are omitted, and the focus is on the substantive assumptions needed for identification. On occasion, some models are a simplified version of the original works to streamline the presentation. Most identification results reported here are expressed in terms of the marginal distribution of one observation and should therefore be applicable to either cross-sectional or time-series settings. The presentation of the estimation methods and their associated asymptotic theory, however, focuses on independent and identically distributed data.

Historically, a large part of the nonlinear measurement error literature has considered the so-called classical measurement error assumptions, which may have two alternative meanings, depending on the context: (*a*) ΔX independent from X^* and $E[\Delta X] = 0$ (which could be called strongly classical) or (*b*) $E[\Delta X | X^* = x^*] = 0$ (which could be called weakly classical, as it allows for heteroskedasticity in the measurement error).

Both conditions are natural generalizations of the standard uncorrelatedness between ΔX and X^* that is typically assumed in linear measurement error models. Depending on the specific model, these two assumptions may include independence from, or conditioning on, some of the other variables of the model (Y, ΔY, Z, ΔZ). Deviations from these assumptions would be labeled nonclassical measurement error, which is currently the focus of intense and increasing attention, as there is well-documented evidence of deviations from classical errors in empirical data (Bollinger 1998, Bound & Krueger 1991, Bound et al. 2001). Nonclassical errors occur, for instance, when there are systematic biases in the errors ΔX, or when random noise does not affect the observed outcome in an additive fashion [e.g., in a so-called nonseparable model (Matzkin 2003)], e.g.,

$$X = a(X^*, \Delta X), \tag{2}$$

where $a(\cdot, \cdot)$ is a general nonlinear function.

This review focuses on nonlinear or nonclassical models (as a linear model with classical measurement errors can be straightforwardly handled via standard instrumental variable techniques). Measurement error models can be classified in terms of the information they use or in terms of the underlying type of model they seek to identify (i.e., the form of the model if the true values of the variables were observed).

The information regarding X^* can take many forms. For example, validation data consist of a separate sample of perfectly observed data that may or may not be matched to (a subset of) individuals in the main data set. Unfortunately, such data are expensive to gather, are often unavailable, and result in estimators whose convergence rate is limited by the typically small size of the validation sample.

Repeated measurements refer to multiple measurements of the same unobserved underlying true unobserved variable X^*, typically contaminated by classical errors. Measurements with Berkson errors also obey a classical-like error structure, but with the role of the measurement and the true value reversed: $X^* = X + \Delta X$. Nonclassical measurements describe any measurements that do not exhibit classical errors. One should, however, distinguish between nonclassical measurements satisfying centering restrictions, which can be handled without validation data (see Hu & Schennach 2008), and general systematic biases, which generally necessitate the use of validation data.

Indicators (or proxies) are related to the true value of the variable interest and may be expressed in different units or may even be nonlinearly related to the true value. They are formally equivalent to measurements with general nonclassical error, but do not necessarily arise from an attempt to measure the true value. Indicators are often monotone in the unobserved variable they proxy for. In a factor model, the indicators may proxy for multiple unobserved variables—i.e., one may not know in advance which indicator provides information regarding which unobserved variable. Instruments, widely used in economic applications, are special cases of indicators, often exhibiting Berkson errors.

Time-series and panel data typically provide either repeated measurements (from future observations) or instruments (from past observations). Measurement error models can also be classified based on the type of underlying model they seek to identify. Examples of underlying models include nonlinear regression models of the form

$$Y = g(X^*, \theta) + \Delta Y, \tag{3}$$

with $E[\Delta Y | X^* = x^*] = 0$ and where $g(X^*, \theta)$ could be nonlinear in both X^* and θ, or could even be a nonparametric function $g(X^*)$.

Another example is maximum likelihood models,

$$\theta = \arg\max_{\theta} E\left[\ln f_{Y,X^*|\theta}(Y, X^*|\theta)\right], \tag{4}$$

where the likelihood function $f_{Y,X^*|\theta}(y, x^*|\theta)$ is specified via a parameter vector θ, whose dimension could increase with sample size to allow for nonparametric or semiparametric sieve estimation (e.g., Gallant & Nychka 1987, Shen 1997, Chen 2007, Hu & Schennach 2008).

Generalized method of moments (GMM) models (Hansen 1982) identify a parameter vector θ via a condition of the form

$$E[m(Y, X^*, \theta)] = 0, \tag{5}$$

where the moment function $m(y, x^*, \theta)$ is parametrically specified. Alternatively, there could be a semiparametric model for the function m within a conditional expectation model ($E[m(Y, X^*, \theta)|Z] = 0$) coupled with sieve estimation (Ai & Chen 2003).

Factor models (e.g., Anderson & Rubin 1956) relate high-dimensional observables Y to the lower-dimensional unobservables X^* of interest and disturbances ΔY:

$$Y = a(X^*, \Delta Y), \tag{6}$$

in which the function $a(\cdot, \cdot)$ may be nonlinear and nonparametric or may be nonseparable in the errors terms ΔY.

More broadly, there is the problem of recovering the unobserved density $f_{X^*}(x^*)$ or, more generally, the unobserved joint density $f_{Y,X^*}(y, x^*)$ of the observable (Y) and unobservable (X^*) variables. From the latter, one can in principle identify any quantity of interest, including the conditional expectation $g(x^*) \equiv E[Y|X^* = x^*]$ and its derivatives. One can even use this approach to combine measurement error problems with other model features, such as endogeneity, and compute appropriately conditioned average effects.

2. THE EFFECT OF MEASUREMENT ERROR

It is well-known that, in a simple linear regression with one regressor, the presence of measurement error causes the estimated slope coefficient to be biased toward zero. This result, known as attenuation, regression to the mean, or the iron law of econometrics (depending on the field), has led to the widespread folklore that the worst that can happen if one neglects the presence of

measurement error is that the regression coefficients merely become less significantly different from zero, so that the resulting statistical inference is conservative, but otherwise valid. However, this optimistic result fails to hold in general for multivariate linear regressions and for nonlinear specifications (Hausman et al. 1995, Hausman 2001). To make matters worse, the standard instrumental variable approach, which is entirely adequate to correct for the endogeneity caused by measurement error in linear models, fails in nonlinear models (Amemiya 1985). These realizations have motivated a large and growing literature that aims to correct for the presence of measurement error in nonlinear models.

The fundamental origin of measurement error bias is that, even under the classical measurement error assumptions, $E[a(X^* + \Delta X)]$ does not equal $E[a(X^*)]$ for a nonlinear function $a(\cdot)$ (e.g., which could be part of the model of interest or an estimator of that model). Chesher (1991) provides a very concise and general way to describe the effect of measurement error in regression models (where the object of interest is $g(x^*) \equiv E[Y|X^* = x^*]$) and in density estimation [where the object of interest is $f_{X^*}(x^*)$]. In the limit of small measurement error variance σ^2, he shows that

$$E[Y|X = x] = g(x) + \frac{\sigma^2}{2}\left(g''(x) + 2g'(x)\left(\ln f_{X^*}(x)\right)'\right) + o\left(\sigma^2\right), \tag{7}$$

$$f_X(x) = f_{X^*}(x) + \frac{\sigma^2}{2} f''_{X^*}(x) + o\left(\sigma^2\right), \tag{8}$$

where primes denote derivatives. It is not fortuitous that these expressions are entirely analogous to the bias of nonparametric kernel estimators with a second-order kernel (Härdle & Linton 1994): The effect of measurement error is to smooth all functions via a convolution with the measurement error density, which plays the role of the kernel. Chesher (1991) also notes that one can use the approximations $g(x) \approx E[Y|X = x]$ and $f_{X^*}(x) \approx f_X(x)$ in Equations 7 and 8 without affecting the order of the $o(\sigma^2)$ remainders. This result is useful because it enables researcher to easily get an idea of what would be the direction and approximate magnitude of bias introduced by a measurement error of a given magnitude. It also gives an intuitive picture of the origin of measurement error bias: Conditional expectations are affected by both the curvature in the regression function [$g''(x)$] and nonuniformities in the density of the regressor [$(\ln f_{X^*}(x))'$], whereas densities are only affected by curvature [$f''_{X^*}(x)$]. Similar results are available for quantile regressions (Chesher 2001), thus implying that the effect on the whole joint distribution of Y and X^* can be assessed. These results have been applied to the estimation of treatment effects with mismeasured covariates (Battistin & Chesher 2014). Extensions to higher-order corrections are not very fruitful because the bias becomes dependent on the specific distribution of the measurement error.

3. METHODS BASED ON DISTRIBUTIONAL INFORMATION

A surprisingly large fraction of the literature on measurement error is devoted to the problem of recovering true error-free quantities from error-contaminated data while assuming that the distribution of the measurement error is known. Although this may not be a very realistic setting, it is a useful setup to discuss some of the basic techniques and difficulties associated with correcting for measurement error.

3.1. Deconvolution

This section begins by introducing the Fourier transform as a very convenient tool to handle classical measurement error. The special case of the Fourier transform of a probability measure is called a characteristic function (CF). The CF of a random vector X taking value in $\mathbb{R}^{d_X}$ is defined

as

$$\phi_X(\xi) \equiv E\left[e^{\mathbf{i}\xi \cdot X}\right] \tag{9}$$

for all $\xi \in \mathbb{R}^{d_X}$ (where $\mathbf{i} \equiv \sqrt{-1}$) and has some remarkable properties (see Lukacs 1970; Loève 1977, sections 13 and 14). For example, there is a one-to-one relationship between $\phi_X(\xi)$ and the probability measure of X. Additionally, $\phi_X(\xi)$ always exists, is continuous, and is bounded (by 1) everywhere. Finally, if $X = X^* + \Delta X$ with X^* and ΔX independent from each other, then one obtains $\phi_X(\xi) = \phi_{X^*}(\xi)\phi_{\Delta X}(\xi)$. This last property, known as the convolution theorem, is particularly useful for classical measurement error problems.

Independence is a sufficient but not necessary condition for the convolution theorem to hold. The necessary and sufficient condition is known as subindependence (Hamedani & Volkmer 2009, Ebrahimi et al. 2010, Schennach 2013a).

If the distribution of the measurement error is known, then $\phi_{\Delta X}(\xi)$ is known, and one can express the CF of the true unobserved variables in terms of known [$\phi_{\Delta X}(\xi)$] or observed [$\phi_X(\xi)$] quantities,

$$\phi_{X^*}(\xi) = \frac{\phi_X(\xi)}{\phi_{\Delta X}(\xi)}, \tag{10}$$

under the standard assumption that $\phi_{\Delta X}(\xi) \neq 0$ [it is sufficient that this holds on a dense subset of $\mathbb{R}$ because $\phi_{X^*}(\xi)$ is continuous]. One can then recover the density of X^* via an inverse Fourier transform:[2]

$$f_{X^*}(x^*) = (2\pi)^{-d_X} \int \phi_{X^*}(\xi)\, e^{-\mathbf{i}\xi \cdot x^*} \mathrm{d}\xi, \tag{11}$$

where integrals without explicit bounds are taken to be over $\mathbb{R}^{d_X}$.

This identification result can be naturally turned into a convenient estimator by taking advantage of the fact that kernel smoothing is also a type of convolution. Specifically, the Fourier transform of a kernel density estimator is given by

$$\hat{\phi}_X(\xi) = \dot{\phi}_X(\xi)\phi_K(h\xi), \tag{12}$$

where

$$\dot{\phi}_X(\xi) = \frac{1}{n}\sum_{j=1}^{n} e^{\mathbf{i}\xi \cdot X_j} \tag{13}$$

is the empirical CF of the sample $(X_1, \ldots, X_n)$, $\phi_K(\xi) = \int K(x)e^{i\xi x}\mathrm{d}x$ is the Fourier transform of the kernel K, and h is the bandwidth. The empirical CF is a common and convenient estimation tool for Fourier-based estimators. It is pointwise root-n consistent, and under mild regularity conditions, it is also uniformly convergent (over an expanding interval) at a slightly less than root-n rate for a compactly supported distribution (Li & Vuong 1998, Hu & Ridder 2010) and, more generally, for distributions satisfying moment existence conditions (Schennach 2004a, Bonhomme & Robin 2010).

This connection with kernel smoothing and deconvolution leads to the kernel deconvolution estimator

$$\hat{f}_{X^*,h}(x^*) \equiv (2\pi)^{-d_X} \int \dot{\phi}_X(\xi) \frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} e^{-\mathbf{i}\xi \cdot x^*} \mathrm{d}\xi, \tag{14}$$

which is the focus of an extensive literature (e.g., Carroll & Hall 1988; Liu & Taylor 1989; Fan 1991a,b; McIntyre & Stefanski 2011).

[2]The result for densities is given for simplicity; a similar result holds for general probability measures (see Loève 1977, section 14.3).

To ensure that the integral in Equation 14 exists, one selects a kernel with compactly supported Fourier transform $\phi_K(\cdot)$ (i.e., an extremely smooth kernel in real space), so that the total amount of noise [in $\dot{\phi}_X(\xi)$] captured by the integral over the support of $\phi_K(h\xi)$ is finite for a given bandwidth h. There is, of course, a variance versus bias trade-off, as can be seen by writing the deviation of the integrand in Equation 14 from its true value $\phi_{X^*}(\xi)$ as

$$\dot{\phi}_X(\xi)\frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} - \phi_{X^*}(\xi) = \left(\dot{\phi}_X(\xi) - \phi_X(\xi)\right)\frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} + \phi_{X^*}(\xi)\left(\phi_K(h\xi) - 1\right). \qquad (15)$$

A small h leads to a small bias [in the second term, $\lim_{h\to 0}\phi_K(h\xi) = \phi_K(0) = 1$ because the kernel K integrates to 1]. But it also leads to a larger support of $\phi_K(h\xi)$ and thus a larger variance (arising from the first term). As sample size n increases, the noise in $\dot{\phi}_X(\xi)$ decreases, and one can afford to gradually decrease h in such a way that both the bias and variance go to zero in order to yield a consistent estimator. This is similar to what happens in a conventional kernel estimator, except that the division by $\phi_{\Delta X}(\xi)$ in Equation 15 results in considerable noise magnification that leads to a slower convergence rate.

More specifically, the rate of convergence of this estimator is governed by the rate of decay of the various Fourier transforms involved as the frequency $|\xi| \to \infty$, which in turn is related to the smoothness of the corresponding original functions. Smoothness in $f_{X^*}(x^*)$ is beneficial, as it reduces the bias (as in a conventional kernel estimator), but smoothness in $f_{\Delta X}(\Delta x)$ is detrimental, as it implies a more rapidly decaying $\phi_{\Delta X}(\xi)$ in the denominator of Equation 15.

In the deconvolution literature, smoothness is traditionally characterized by bounds on the rate of decay of the tail of a CF. There are two broad typical behaviors. First, there are densities that are ordinarily smooth, whose Fourier transforms have a tail bounded above and below by a multiple of $|\xi|^{-\alpha}$ for some $\alpha > 0$. These essentially correspond to densities admitting a finite number of derivatives (see, e.g., Schennach 2004a, theorem 3, for an upper bound on the CF and Hu & Ridder 2010, theorem 1, for a lower bound).[3] Second, there are densities that are supersmooth, whose Fourier transforms have a tail bounded above and below by some multiple of $|\xi|^{\alpha}\exp(-|\xi|^{\beta}/\gamma)$ for some $\alpha, \beta, \gamma > 0$. These essentially correspond to densities that are infinitely many times differentiable (see Schennach 2007a, lemma S.2).

Fan (1991b) derives bounds on the convergence rates in density deconvolution problems, revealing that, when the densities of both X^* and ΔX are ordinarily smooth, the best possible (root mean square) convergence rate is of the form $n^{-\eta}$ for some $\eta > 0$. In the extreme nonsmooth case in which ΔX contains a mixture with a single point mass, An & Hu (2012) observe that very fast convergence rates ($\eta = -2/5$) can be obtained. This case has empirical relevance in self-reported data when individuals may have a finite probability of reporting the truth exactly.

When the density of ΔX is supersmooth, although the density of X^* is still ordinarily smooth, the optimal convergence rate becomes a mere $(\ln n)^{-\eta}$ for some $\eta > 0$, which is very slow. However, this result is somewhat artificially pessimistic because it makes asymmetric assumptions regarding $\phi_X(\xi)$ (which is at best assumed ordinarily smooth) and $\phi_{\Delta X}(\xi)$ (which could be supersmooth). If $\phi_X(\xi)$ is also allowed to be supersmooth, then fast convergence rates are again possible even for supersmooth errors (as shown in Schennach 2004c in a more general context). Moreover, fast convergence rates are also possible in many semiparametric settings (see Taupin 2001, Schennach 2004a, Hu & Ridder 2010, Schennach et al. 2012), with either ordinary smooth or supersmooth measurement error distributions.

[3] Hu & Ridder (2010) show that compact support implies ordinary smoothness under some asymmetry and nonsmoothness conditions at the boundaries of the compact support.

The idea of kernel deconvolution can be extended to cover nonparametric regression as well (Fan & Truong 1993):

$$\hat{g}_b(x^*) = \frac{\int \frac{\hat{\phi}_X^Y(\xi)}{\phi_{\Delta X}(\xi)} \phi_K(h\xi)\, e^{-\mathbf{i}\xi\cdot x^*}\, \mathrm{d}\xi}{\int \frac{\hat{\phi}_X(\xi)}{\phi_{\Delta X}(\xi)} \phi_K(h\xi)\, e^{-\mathbf{i}\xi\cdot x^*}\, \mathrm{d}\xi}, \tag{16}$$

where

$$\hat{\phi}_X^Y(\xi) \equiv \sum_{j=1}^{n} Y_j e^{\mathbf{i}\xi\cdot X_j}, \tag{17}$$

which is a quantity related to the empirical CF and which can be viewed as an estimator of $E[Ye^{\mathbf{i}\xi\cdot X}]$ or, equivalently, of $\int E[Y|X=x]e^{\mathbf{i}\xi\cdot x}\mathrm{d}F(x)$ or $[\partial\phi_{YX}(\zeta,\xi)/\partial\zeta]_{\zeta=0}$. It exhibits convergence properties similar to the empirical CF (Schennach 2004a, Bonhomme & Robin 2010).

3.2. Validation Data

For a main sample containing data on Y and X (where X is mismeasured), validation data (e.g., Sepanski & Carroll 1993) typically consist of an auxiliary sample containing data on both X and X^* that can be used to straightforwardly recover the distribution of the measurement error or the density $f_{X^*|X}(x^*|x)$. (More general forms of data combination between samples are discussed in Ridder & Moffitt 2007.) Under the assumptions that $E[Y|X=x, X^*=x^*] = E[Y|X^*=x^*]$ and that $f_{X^*|X}(x^*|x)$ is transferable across the two samples, it is straightforward to correct for measurement error in a parametric regression setting (Equation 3), via the equality

$$E[Y|X] = \int E[Y|X=x, X^*=x^*]\, f_{X^*|X}(x^*|x)\, \mathrm{d}x^* = \int E[Y|X^*=x^*]\, f_{X^*|X}(x^*|x)\, \mathrm{d}x^*. \tag{18}$$

Indeed, one merely needs to define a modified regression function in terms of the original specification $g(x^*, \theta)$,

$$\tilde{g}(x,\theta) \equiv \int g(x^*,\theta)\, f_{X^*|X}(x^*|x)\, \mathrm{d}x^*, \tag{19}$$

where $f_{X^*|X}(x^*|x)$ can be estimated from the validation sample. Then, a conventional least-square projection of Y on X using the model

$$Y = \tilde{g}(X,\theta) + \Delta Y \tag{20}$$

identifies θ and leads to a natural estimator (Sepanski & Carroll 1993). A similar reasoning can, of course, be used to handle general GMM-based models (Chen et al. 2005).

The main advantage of validation data is that they offer a way to handle measurement when it is not of a classical nature. Unfortunately, the availability of validation data is the exception rather than the rule. In economics, Bound & Krueger (1991) provide one widely cited validation data set, but it is difficult to find many other examples of true validation data. For this reason, methods to handle nonclassical measurement error without relying on validation data are being developed (see Section 6).

Hu & Ridder (2012) suggest an interesting setup in the context of classical measurement error, in which the main sample contains Y and X, whereas the validation sample contains only X^* (instead of matched observations on X and X^*). The two samples are assumed to be drawn from the same population. Taking the regression case as an example, their idea can summarized as follows: The measurement error distribution is obtained by deconvolving the distribution of X from the main sample by the distribution of X^* from the validation sample. Then, identification

of the regression function $E[Y|X^* = x^*]$ follows from deconvolution arguments (in the spirit of Equation 16).

4. METHODS BASED ON AUXILIARY VARIABLES

4.1. Repeated Measurements

Repeated measurements X and Z are related to the true underlying variable of interest X^* via

$$X = X^* + \Delta X, \tag{21}$$

$$Z = X^* + \Delta Z, \tag{22}$$

where the appropriate assumptions regarding the measurement errors ΔX and ΔZ are specified below. Repeated measurements are commonly available in data sets when the same survey or test is repeated over time or if the same question is asked to different people (e.g., spouses, employer/employee).

Repeated measurements are useful because the distribution of the true unobserved variable X^* can be obtained via an old but very powerful result known as Kotlarski's lemma (see Kotlarski 1967; Rao 1992, p. 21). This lemma (proven more generally below) states that if X and Z take value in $\mathbb{R}$ and X^*, with ΔX and ΔZ mutually independent with $E[\Delta X] = 0$, then (provided $E[e^{\mathbf{i}\zeta Z}] \neq 0$ for all real ζ),

$$\phi_{X^*}(\xi) = \exp\left(\int_0^{\xi} \frac{E\left[\mathbf{i}Xe^{\mathbf{i}\zeta Z}\right]}{E\left[e^{\mathbf{i}\zeta Z}\right]} \mathrm{d}\zeta\right). \tag{23}$$

Kotlarski's lemma has been modified and generalized in various ways since then, as outlined below.

In the very special case in which the distributions of ΔX and ΔZ are identical and symmetric about zero, one can obtain the CF of X^* in a simpler way (see Horowitz & Markatou 1996, Li & Vuong 1998, Delaigle et al. 2008) by noting that $\phi_{Z-X}(\xi) = \phi_{\Delta Z-\Delta X}(\xi) = \phi_{\Delta X}(\xi)\phi_{\Delta X}(-\xi) = |\phi_{\Delta X}(\xi)|^2$ so that $\phi_{X^*}(\xi) = \phi_X(\xi)/\sqrt{|\phi_{Z-X}(\xi)|}$ (selecting the positive root). An alternative expression (Delaigle et al. 2008) that does not require the distributions of ΔX and ΔZ to be identical is $\phi_{X^*}(\xi) = \phi_{(X+Z)/2}(\xi)/\phi_{(X-Z)/2}(\xi)$.

Li & Vuong (1998) prove the consistency of a nonparametric density estimator based on Kotlarski's lemma. This result is extended by Li (2002) to multivariate settings[4] (assuming mutual independence of the elements of both ΔX and ΔZ) and used to correct for measurement error in a nonlinear regression model (Equation 3). The idea is to first use the following identity:

$$\tilde{g}(x,\theta) \equiv E[Y|X = x] = \int g(x^*,\theta) f_{X^*|X}(x^*|x)\,\mathrm{d}x^* = \int g(x^*,\theta) \frac{f_{\Delta X}(x - x^*) f_{X^*}(x^*)}{f_X(x)}\mathrm{d}x^*, \tag{24}$$

where the densities $f_{\Delta X}$ and f_{X^*} can be estimated nonparametrically consistently using repeated measurements. This can be viewed as a semiparametric extension of the fully parametric treatment of Hsiao (1989). The parameter θ can then be estimated by minimizing the sample analog of $E[(Y - \tilde{g}(X,\theta))^2]$.

[4]Both Li & Vuong (1998) and Li (2002) require the observed variable Z to have a nonvanishing CF and a compactly supported distribution. These requirements are not mutually exclusive: For instance, take a mixture of two centered symmetric triangular distributions with widths w_1 and w_2 such that w_1/w_2 is irrational.

Hausman et al. (1991, 1995) take a different approach to identification and estimation, by focusing on moments of X^*. Hausman et al. (1991) consider a polynomial model of the form

$$Y = \sum_{j=0}^{J} \theta_j (X^*)^j + \Delta Y, \tag{25}$$

where X^* and Y are scalars, and the disturbance satisfies weaker requirements than for Kotlarski's identity: $E[\Delta Y | X^*, \Delta Z] = 0$, $E[\Delta X | X^*, \Delta Z] = 0$, and ΔZ independent from X^*. Then they note that if X^* were observed, $\theta \equiv (\theta_0, \ldots, \theta_J)$ would be identified from

$$\theta = \left(E\left[VV'\right]\right)^{-1} (E[VY]), \qquad V = \left(1, X^*, \ldots, (X^*)^J\right)', \tag{26}$$

where the elements of $E[VV']$ have the general form $\zeta_m = E[(X^*)^m]$, $m = 0, \ldots, 2J$, and elements of $E[VY]$ have the general form $\xi_l = E[Y (X^*)^l]$, $l = 0, \ldots, J$. Hence, one needs to identify ξ_l and ζ_m from observed moments. By computing moments of the form $E[XZ^j]$, $E[Z^j]$, and $E[YZ^j]$ in terms of the unobservable moments ζ_j, ξ_j, and $\nu_j = E[(\Delta Z)^j]$, they arrive at the following relationships:

$$\zeta_j = E[XZ^{j-1}] - \sum_{l=0}^{j-2} \binom{j-1}{l} \zeta_{l+1} \nu_{j-l-1}, \tag{27}$$

$$\nu_j = E\left[Z^j\right] - \sum_{l=1}^{j} \binom{j}{l} \zeta_l \nu_{j-l}, \tag{28}$$

$$\xi_j = E\left[YZ^j\right] - \sum_{l=0}^{j-1} \binom{j}{l} \xi_l \nu_{j-l}. \tag{29}$$

Starting from $\nu_0 = 1$ and $\zeta_1 = E[X]$, the relations in Equations 27 and 28 can be used to recursively identify ν_1, ζ_2, ν_2, $\zeta_3, \ldots,$ ν_{2J-1}, and ζ_{2J}. Then, $\xi_0, \ldots, \xi_J$ are obtained from Equation 29.

Hausman et al. (1995) go a step further and observe that, because polynomials form a basis for functions in L_2 (using a suitable weighted L_2 norm), the polynomial result can be used to identify general nonlinear regression models via a two-step process:

1. Use the polynomial result with repeated measurement to obtain $\hat{g}(x^*)$, a nonparametric series estimate of $E[Y | X^* = x^*]$.
2. Solve
$$\hat{\theta} = \arg\min_{\theta} \int \left(g(x^*, \theta) - \hat{g}(x^*)\right)^2 w(x^*) \, \mathrm{d}x^* \tag{30}$$
for some given weighting function $w(x^*)$.

In the case of general nonlinear regressions, both Li (2002) and Hausman et al. (1995) only establish consistency (without a rate or limiting distribution) and do so under rather stringent moment existence conditions (compact support in the case of Li 2002 and moment-generating function existence in the case of Hausman et al. 1995).

Schennach (2004a) obtains an identity (which is recast here in a slightly different form and under slightly different assumptions for expository purposes) that generalizes both Kotlarski's result and Hausman et al. (1991). Because a large class of models can be identified from the knowledge of a set of moments, the result is phrased in a form that directly identifies moments involving (a possibly multivariate) X^* (and perhaps other perfectly measured variables). This is true by construction for GMM-type models, and also for likelihood models. As a special case, for nonlinear regression

models (Equation 3), the moment vector needed is $E[(Y - g(X^*, \theta))\partial g(X^*, \theta)/\partial\theta]$ (through the first-order conditions for least-square minimization).

Repeated measurements can be used to identify such moments (in which X^*, X, Y, and Z could be random vectors, with $d_{X^*} = d_X = d_Z$). To illustrate, if (*a*) $E[\Delta X \,|X^*, \Delta Z] = 0$, (*b*) $E[Y \,|X^*, \Delta Z] = E[Y \,|X^*]$,[5] and (*c*) ΔZ is independent from X^*, then, for any function $u(x^*)$ with Fourier transform $\mu(\xi)$, one has

$$E\left[Yu\left(X^*\right)\right] = (2\pi)^{-d_X} \int_{\mathbb{R}^{d_X}} \mu\left(-\xi\right) \frac{E\left[Ye^{\mathbf{i}\xi\cdot Z}\right]}{E\left[e^{\mathbf{i}\xi\cdot Z}\right]} \phi_{X^*}\left(\xi\right) \mathrm{d}\xi, \tag{31}$$

where the integral is over the whole $\mathbb{R}^{d_X}$ space, and the CF of X^* is given by

$$\phi_{X^*}\left(\xi\right) = \exp\left(\int_0^{\xi} \frac{E\left[\mathbf{i}Xe^{\mathbf{i}\zeta\cdot Z}\right]}{E\left[e^{\mathbf{i}\zeta\cdot Z}\right]} \cdot \mathrm{d}\zeta\right), \tag{32}$$

where the integral is the path integral of a vector-valued field along a piecewise smooth path joining the origin and the point $\xi \in \mathbb{R}^{d_X}$ (provided all the requisite quantities exist and the denominators are nonvanishing).

It is instructive to outline the proof of this result.[6] The ratio in Equation 32 can be written as

$$\begin{aligned}
\frac{E\left[\mathbf{i}Xe^{\mathbf{i}\zeta\cdot Z}\right]}{E\left[e^{\mathbf{i}\zeta\cdot Z}\right]} &= \frac{E\left[\mathbf{i}X^*e^{\mathbf{i}\zeta\cdot(X^*+\Delta Z)}\right] + E\left[\mathbf{i}\Delta Xe^{\mathbf{i}\zeta\cdot(X^*+\Delta Z)}\right]}{E\left[e^{\mathbf{i}\zeta\cdot(X^*+\Delta Z)}\right]} \\
&= \frac{E\left[\mathbf{i}X^*e^{\mathbf{i}\zeta\cdot(X^*+\Delta Z)}\right] + E\left[\mathbf{i}E\left[\Delta X \,|X^*, \Delta Z\right]e^{\mathbf{i}\zeta\cdot(X^*+\Delta Z)}\right]}{E\left[e^{\mathbf{i}\zeta\cdot(X^*+\Delta Z)}\right]} \\
&= \frac{E\left[\mathbf{i}X^*e^{\mathbf{i}\zeta\cdot X^*}\right]E\left[e^{\mathbf{i}\zeta\cdot\Delta Z}\right]}{E\left[e^{\mathbf{i}\zeta\cdot X^*}\right]\left[e^{\mathbf{i}\zeta\cdot\Delta Z}\right]} = \frac{E\left[\mathbf{i}X^*e^{\mathbf{i}\zeta\cdot X^*}\right]}{E\left[e^{\mathbf{i}\zeta\cdot X^*}\right]} = \nabla_\zeta \ln E\left[e^{\mathbf{i}\zeta\cdot X^*}\right],
\end{aligned} \tag{33}$$

where the definition of the repeated measurements, iterated expectations, the conditional mean and independence assumptions regarding the errors (ΔX and ΔZ, respectively), various cancelations, and the chain rule have been used in turn. Next, Equation 32 can be shown by noting that the path integral of a gradient yields the original function:

$$\exp\left(\int_0^{\xi} \nabla_\zeta \ln E\left[e^{\mathbf{i}\zeta\cdot X^*}\right]\cdot \mathrm{d}\zeta\right) = \exp\left(\ln E\left[e^{\mathbf{i}\xi\cdot X^*}\right] - \ln E\left[e^{\mathbf{i}0\cdot X^*}\right]\right) = E\left[e^{\mathbf{i}\xi\cdot X^*}\right] = \phi_{X^*}\left(\xi\right). \tag{34}$$

The integrand in Equation 31 can then be written as

$$\begin{aligned}
\frac{E\left[Ye^{\mathbf{i}\xi\cdot Z}\right]}{E\left[e^{\mathbf{i}\xi\cdot Z}\right]}E\left[e^{\mathbf{i}\xi\cdot X^*}\right] &= \frac{E\left[E\left[Y \,|X^*, \Delta Z\right]e^{\mathbf{i}\xi\cdot(X^*+\Delta Z)}\right]}{E\left[e^{\mathbf{i}\xi\cdot(X^*+\Delta Z)}\right]}E\left[e^{\mathbf{i}\xi\cdot X^*}\right] \\
&= \frac{E\left[E\left[Y \,|X^*\right]e^{\mathbf{i}\xi\cdot X^*}\right]E\left[e^{\mathbf{i}\xi\cdot\Delta Z}\right]}{E\left[e^{\mathbf{i}\xi\cdot X^*}\right]E\left[e^{\mathbf{i}\xi\cdot\Delta Z}\right]}E\left[e^{\mathbf{i}\xi\cdot X^*}\right] = E\left[E\left[Y \,|X^*\right]e^{\mathbf{i}\xi\cdot X^*}\right] \\
&= \int E\left[Y \,|X^* = x^*\right] f_{X^*}\left(x^*\right) e^{i\xi\cdot x^*}\mathrm{d}x^* \equiv \alpha\left(\xi\right),
\end{aligned} \tag{35}$$

where iterated expectations, the conditional independence assumption regarding Y and ΔZ, and various cancelations have been used in turn. Equation 31 is then obtained by using Parseval's identity $(2\pi)^{-d_X}\int \mu(-\xi)\alpha(\xi)\mathrm{d}\xi = \int u(x^*)a(x^*)\mathrm{d}x^*$, where $a(x^*) = E[Y \,|X^* = x^*]f_{X^*}(x^*)$ with Fourier transform $\alpha(\xi)$.

[5] In a regression setting, this is implied by $E[\Delta Y \,|X^*, \Delta Z] = 0$.

[6] This is done without dwelling on technical issues such as the interchange of integrals, derivatives and expectations, etc.

A few remarks are in order. Perfectly measured variables (or variables contaminated by an error with zero mean conditional on X^*, such as the dependent variable in a regression[7]) can be included in Y if they enter the moments linearly. If Y is not needed, it can be simply set to $Y = 1$. If the model is such that a zero-mean measurement error in the dependent variable could introduce bias (e.g., if the dependent variable enters nonlinearly in the moment conditions of the GMM or in the first-order conditions for maximum likelihood), then the dependent variable should be considered as part of the X^* vector (rather than part of Y) in the above formalism. Perfectly measured variables that enter the moment nonlinearly can be included in X^*: The corresponding elements of X and Z can then just be set to be equal. Alternatively, perfectly measured variables can be handled under weaker conditions by expanding the moment function into a hybrid basis (a Fourier basis for the mismeasured variables and a general basis for the correctly measured variables), as shown in Schennach (2004a).

The Fourier transform $\mu(\xi)$ may be a generalized function (Lighthill 1962, Temple 1963, Gel'fand & Shilov 1964, Schwartz 1966). For instance, if $u(x^*)$ is a polynomial, the Fourier transform $\mu(\xi)$ consists of delta function derivatives of various orders that effectively extract various derivatives of the quantity $(E[Ye^{\mathbf{i}\xi\cdot Z}]/E[e^{\mathbf{i}\xi\cdot Z}])\phi_{X^*}(\xi)$ in Equation 31. In this fashion, one can recover the polynomial result of Hausman et al. (1991) under the same conditional mean and independence assumptions.

Equations 31 and 32 also suggest a very natural estimator in which all quantities of the form $E[Ve^{\mathbf{i}\zeta\cdot Z}]$ for $V = 1, Y$ are replaced by sample averages $n^{-1}\sum_{j=1}^{n} V_j e^{\mathbf{i}\zeta\cdot Z_j}$. After a simple automatic bounding device (ensuring, e.g., that an estimated CF obtained via Equation 32 is bounded), Schennach (2004a) shows that this approach yields a root-n consistent and asymptotically normal estimator that does not require any user-specified bandwidth parameter, provided $u(x^*)$ is sufficiently smooth. This smoothness condition ensures that $\mu(\xi)$ decays sufficiently rapidly as $|\xi| \to \infty$ to downweigh the noise in the tail in the estimated CF, yielding a finite overall noise that decays to zero at the rate $n^{-1/2}$ as $n \to \infty$. Schennach et al. (2012) extend these results to general semiparametric functionals of densities, conditional expectations, and derivatives thereof.

One can also recover Kotlarski's identity from Equations 31 and 32 under fewer independence assumptions by setting $Y = 1$ and $\mu(\xi) = e^{\mathbf{i}\xi\cdot x_0^*}$ [which corresponds to setting $u(x^*)$ to be a delta function, or a point mass, at x_0^*]. In a similar vein, only setting $\mu(\xi) = e^{\mathbf{i}\xi\cdot x_0^*}$ (but keeping Y) yields the identification of $E[Y|X^* = x_0^*]f_{X^*}(x_0^*)$, which opens the way to nonparametric identification of conditional expectations [after division by $f_{X^*}(x_0^*)$, which is also identified].

In this nonparametric setting, an estimator cannot be obtained by merely replacing all expectations of the form $E[Ve^{\mathbf{i}\zeta\cdot Z}]$ by the corresponding sample average in Equations 31 and 32, because this yields noise of infinite magnitude, as $\mu(\xi)$ does not decay as $|\xi| \to \infty$. As shown in Schennach (2004c), this problem can be solved via smoothing by a symmetric kernel $K(\cdot)$ of width h, by setting $u(x^*) = h^{-1}K(h^{-1}(x^* - x_0^*))$. This corresponds to $\mu(\xi) = \phi_K(h\xi)e^{\mathbf{i}\xi\cdot x_0^*}$, where the dependence of h is suppressed in the notation. In analogy with kernel deconvolution, the kernel is selected so that its Fourier transform $\phi_K(\xi)$ has compact support. This ensures that the integrated noise is finite, so that a consistent estimator of $g(x_0^*) \equiv E[Y|X^* = x_0^*]$, denoted $\hat{g}_h(x_0^*)$, can be obtained by replacing $E[Ve^{\mathbf{i}\xi\cdot Z}]$ for $V = 1, X, Y$ by $n^{-1}\sum_{j=1}^{n} V_j e^{\mathbf{i}\xi\cdot Z_j}$ in

$$\frac{\int \phi_K(h\xi)\, e^{-\mathbf{i}\xi\cdot x_0^*} \frac{E[Ye^{\mathbf{i}\xi\cdot Z}]}{E[e^{\mathbf{i}\xi\cdot Z}]} \phi_{X^*}(\xi)\, \mathrm{d}\xi}{\int \phi_K(h\xi)\, e^{-\mathbf{i}\xi\cdot x_0^*} \phi_{X^*}(\xi)\, \mathrm{d}\xi}, \tag{36}$$

[7]This is not possible in the case of a dependent variable contaminated by nonclassical measurement error, a case considered in Section 6.

with $\phi_{X^*}(\xi)$ given by Equation 32, if one lets $h \to 0$ at a suitable rate as $n \to \infty$. Schennach (2004c) derives the convergence rate of this estimator as a function of the smoothness of the densities and conditional expectations involved (via the rate of decay of their Fourier transforms). An important finding is that the convergence rates are often comparable to the case in which the distribution of the measurement error is fully known. These results are generalized in Schennach et al. (2012) to yield nonparametric estimates of densities and conditional expectations (and derivatives thereof) that are uniformly consistent (in some cases over expanding intervals).

Kotlarski's identity has been generalized in other ways. For instance, to relax some of the nonvanishing CF assumptions, Evdokimov (2009) assumes mutual independence of ΔX, ΔZ, and X^* and suggests using $Z - X$ and Z as repeated measurements of ΔZ to identify its distribution, from which one can recover the distribution of X^* via standard deconvolution of the distribution of Z by the distribution of ΔZ. This only requires ΔZ to have a nonvanishing CF but not X^*. This can be even further relaxed by evaluating an appropriate limit, provided that higher-order derivatives of the CF do not vanish where the CF itself does. This idea is exploited in Schennach (2000) and Evdokimov & White (2012). Also, the structure of the second measurement can be relaxed to $Z = a + bX^* + \Delta Z$ with $E[\Delta Z] = 0$ if another variable, Y, related to X^* but independent from the errors, is available. Carroll et al. (2004) observe that the slope coefficient can be identified from $b = \mathrm{Cov}(Y, Z)/\mathrm{Cov}(Y, X)$, whereas the intercept is given by $a = E[Z] - bE[X]$. A more general polynomial measurement structure is considered by Hu & Sasaki (2015), as discussed in Section 6.2.

4.2. Factor Models

It is also possible to extend Equations 21 and 22 to a more general factor model structure between the observed measurements X and the unobserved factors X^*:

$$X = \Lambda X^* + \Delta X, \tag{37}$$

where Λ is a $d_X \times d_{X^*}$ matrix of factor loadings (here $d_X \neq d_{X^*}$ in general, so in this notation, the repeated measurement Z is superfluous), where the errors ΔX are independent from X^* and the elements of the vector ΔX are typically assumed mutually independent (although these restrictions can be relaxed).

The factor loadings can be identified, up to some normalizations, from the covariance matrix of X (Anderson & Rubin 1956) if the matrix Λ is such that there remain two disjoint matrices of rank d_{X^*} after any one row of Λ is removed. Normalizations are necessary because it is possible to substitute $\Lambda = \tilde{\Lambda} T^{-1}$ and $X^* = T\tilde{X}^*$ into Equation 37, where T is any invertible $d_{X^*} \times d_{X^*}$ matrix, without affecting the observable quantities, while obtaining different factors $\tilde{X}^*$ and different factor loadings $\tilde{\Lambda}$. A popular normalization is to assume that the elements of X^* are mutually independent with unit variance (and are ordered according to the fraction of the variance of the observed quantities they explain). An alternative normalization is to require some of the observed measurements to be dedicated to a given element of the unobservable X^*, in which case the elements of X^* can be generally correlated and some of the independence assumptions regarding X^* and ΔX can be relaxed (Cunha et al. 2010). Heckman et al. (2010a) observe that, when the latent factors X^* are used as control variables to match comparable individuals in the analysis of treatment effects, the need for normalizations is entirely eliminated. While identification of Λ from the covariances alone clearly requires $d_X \geq 2d_{X^*} + 1$ to satisfy the requisite rank conditions, that number can be reduced using higher-order moments (Bonhomme & Robin 2009) or, equivalently, higher-order derivatives of the log CF (Ben-Moshe 2013), as discussed in Section 5.1.

Once the factor loading matrix Λ is known, one can construct two vectors of repeated measurements suitable for use in Equations 31 and 32 to estimate the joint distribution of the factors X^* (Heckman et al. 2010b). Specifically, assume that one can decompose Λ' as $(\Lambda'_A, \Lambda'_B, \Lambda'_C)$ where Λ_A and Λ_B are $d_{X^*} \times d_{X^*}$ invertible submatrices and Λ_C has any dimension (it even can be empty, as it plays no role in the identification of the distribution of X^*), and let $X' = (X'_A, X'_B, X'_C)$ be the corresponding partitioning of X in Equation 37. Then, the repeated measurement vectors X and Z in Equations 31 and 32 can be taken to be $X \equiv \Lambda_A^{-1} X_A$ and $Z \equiv \Lambda_B^{-1} X_B$. If the elements of the vector X^* are assumed mutually independent, Bonhomme & Robin (2010) observe that it is possible to recover the distribution of X^* using fewer measurements. To exploit these additional independence assumptions, they employ a Kotlarski-type identity that relies on second-order derivatives instead of the first-order derivatives corresponding to $E[\mathbf{i}Xe^{\mathbf{i}\zeta Z}]$ in Kotlarski's identity (Equation 23). Ben-Moshe (2014) provides a range of identification theorems allowing for various levels of dependence between the factors, from general dependence to mean independence. His results, many of which are necessary and sufficient, are phrased in terms of first or second derivatives of the CFs. Interestingly, some of the rank conditions are reminiscent of the definition of the Kruskal (1977) rank, discussed in Section 6.1.

A generalization (under suitable normalizations) to fully nonlinear nonseparable factor models, i.e.,

$$X = g(X^*, \Delta X),$$

for some nonlinear function $g(\cdot, \cdot)$, is developed in Cunha et al. (2010), using and extending some of the techniques found in Hu & Schennach (2008) (discussed in more detail in Section 6.2).

In summary, repeated measurements and their factor model extension provide a very powerful and general approach to the identification and estimation of measurement error models that does not require knowledge of the measurement error distribution. In fact, it has recently been suggested that surveys should be designed to elicit multiple measurements that may be mismeasured rather than attempting to gather exact data (Browning & Crossley 2009).

4.3. Instrumental Variables

Instrumental variables typically consist of a vector Z of random variables (a) that are related in some way to the true unobserved variable X^*, for instance, through some relationship of the form

$$X^* = h(Z) + \Delta Z, \tag{38}$$

where $h(Z)$ is some linear or nonlinear function, and (b) that satisfy some exclusion restrictions, such as $E[\Delta X | Z] = 0$ or $E[\Delta Y | Z] = 0$.

Instruments are more general than repeated measurements in the sense that they can be biased indicators of the true unobserved X^*. The precise relationship between X^* and Z and the specific notion of exclusion may depend on the approach considered.

Hausman et al. (1991) consider a general polynomial model with scalar Y and X^* and with an instrument equation of the form of Equation 38 where $E[\Delta Y | Z, \Delta Z] = 0$, $E[\Delta X | Z, \Delta Z, \Delta Y] = 0$, $E[\Delta Z] = 0$, and ΔZ is independent from Z. The function $h(\cdot)$ in Equation 38 is identified from $E[X | Z] = E[X^* | Z] = h(Z)$. In fact, for identification purposes, one can simply use $h(Z)$ as the instrument and hence relabel $h(Z)$ as simply Z in the sequel. Under these assumptions, the following observable conditional expectations can be shown to have a polynomial form:

$$E[Y | Z = z] = \sum_{j=0}^{J} \gamma_j z^j, \tag{39}$$

$$E[XY|Z=z] = \sum_{j=0}^{J+1} \beta_j z^j. \tag{40}$$

The identified coefficients can then be used in a recursive relation to identify θ. Start with $v_0 = 1$, $v_1 = 0$, $\theta_J = \gamma_J$, and $\theta_{J-1} = \gamma_{J-1}$. For $k > 1$, one then uses the fact that

$$v_j = \left(\binom{J}{J-j+1}\theta_J\right)^{-1}\left(\beta_{J-j+1} - \gamma_{J-j} - \sum_{l=J-j+1}^{J-1}\binom{l}{J-j+1}\theta_l v_{l-J+j}\right), \tag{41}$$

$$\theta_{J-j} = \gamma_{J-j} - \sum_{l=J-j+1}^{J}\binom{l}{J-j}\theta_l v_{l-J+j}. \tag{42}$$

These equations can be used to recursively find $v_2, \theta_{J-2}, v_3, \theta_{J-3}, \ldots, v_J$, and θ_0. Interestingly, unlike the repeated measurement case, this solution does not directly generalize to arbitrary nonlinear or nonparametric models because the recursive relationships would have to start the recursion from moments of infinite order.

To avoid this problem, Newey (2001) extends the polynomial to general nonlinear parametric models of the form

$$Y = g(X^*, \theta) + \Delta Y \tag{43}$$

by starting from the following integral equations implied by the same assumptions as the polynomial model above:

$$E[Y|Z=z] = \int g(z-u,\theta)\,\mathrm{d}F_U(u), \tag{44}$$

$$E[XY|Z=z] = \int (z-u)\,g(z-u,\theta)\,\mathrm{d}F_U(u), \tag{45}$$

where F_U is the distribution of $U \equiv -\Delta Z$. For estimation purposes, he uses a vector of unconditional moment conditions implied by these relationships in which $\mathrm{d}F_U(u)$, the distribution of U, is modeled via a nonparametric sieve. Although this work did not include a formal proof of identification of the model from Equations 44 and 45, it did lay down the basic equations that later led to a proof of identification.

In fact, proving identification is not a trivial task in this model: In particular, Schennach (2004b) shows that a class of exponential specifications is in fact not identified from Equations 44 and 45 alone, thereby pointing out cases that a complete identification result must carefully exclude. The proof is based on the fact that exponentials are shape invariant under a convolution, so that the effect of measurement error cannot be distinguished from a change in the prefactor of the exponential.

In a specification $g(x^*, \theta)$ with a d_{X^*}-dimensional x^*, Wang & Hsiao (2011) use Equations 44 and 45 to show how to identify θ if it contains at most $d_{X^*} + 1$ elements (which is unfortunately no more than the number of parameters a linear specification would have). Their identification strategy rests on the strong assumption that $\int |g(x^*, \theta)|\mathrm{d}x^* < \infty$, although they argue that a combination of truncation and limiting arguments should imply that identification holds more generally.

Schennach (2007a) provides a general proof of nonparametric identification based on the Fourier transforms of Equations 44 and 45:

$$\varepsilon_Y(\zeta) = \gamma(\zeta)\,\phi(\zeta), \tag{46}$$

$$\mathbf{i}\varepsilon_{XY}(\zeta) = \dot{\gamma}(\zeta)\,\phi(\zeta), \tag{47}$$

where the overdot denotes derivatives and, with $V = Y$ or $V = XY$,

$$\varepsilon_V(\zeta) = \int E[V|Z = z]\, e^{\mathbf{i}\zeta z} \mathrm{d}z, \tag{48}$$

$$\gamma(\zeta) = \int g(x^*)\, e^{\mathbf{i}\zeta x^*} \mathrm{d}x^*, \tag{49}$$

$$\phi(\zeta) = E\left[e^{\mathbf{i}\zeta U}\right] = E\left[e^{-\mathbf{i}\zeta \Delta Z}\right]. \tag{50}$$

These identities can then be manipulated to secure nonparametric identification of $g(x^*)$, under the following conditions: (*a*) $E[|\Delta Z|] < \infty$, whereas $|g(x^*)|$, $|E[Y|Z = z]|$, and $|E[XY|Z = z]|$ are bounded by polynomials, and (*b*) $\phi(\zeta)$ is nonvanishing and $\gamma_o(\zeta)$ is nonvanishing almost everywhere,[8] where $\gamma_o(\zeta)$ denotes the ordinary function component[9] of $\gamma(\zeta)$. This allows for the fact that the Fourier transform of a function bounded by a polynomial, but not absolutely integrable, such as most conditional expectations, may not be a function in the usual sense but a tempered distribution, a class of well-behaved generalized functions (Lighthill 1962, Temple 1963, Gel'fand & Shilov 1964, Schwartz 1966). The regression function is then given by[10]

$$g(x^*) = (2\pi)^{-1} \int \frac{\varepsilon_Y(\zeta)}{\phi(\zeta)} e^{-\mathbf{i}\zeta x^*} \mathrm{d}\zeta, \tag{51}$$

where

$$\phi(\zeta) = \exp\left(\int_0^\zeta \frac{\mathbf{i}\varepsilon_{(Z-X)Y,o}(\xi)}{\varepsilon_{Y,o}(\xi)} \mathrm{d}\xi\right), \tag{52}$$

and where, for $V = Y$ or $(Z - X)Y$, $\varepsilon_{V,o}(\xi)$ denotes the ordinary function component of $\varepsilon_V(\xi)$. Schennach (2007a) also provides a root-n consistent estimator [when $g(x^*)$—but not the distribution of ΔZ—is parametrically specified] based on moment conditions implied by Equations 46 and 47.

Although there is some similarity between Equation 52 and Kotlarski's result (Equation 23), there are important differences. First, Equation 52 involves the Fourier transforms of conditional expectations rather than probability densities. Second, the relationships between all variables Y, X, and Z play a role in Equation 52, whereas only X and Z enter Equation 23.

Nadai & Lewbel (2016) push this approach further by (*a*) adapting it to handle some forms of correlations between the errors in the dependent variable Y and the regressor X, (*b*) allowing for multiplicative errors, and (*c*) providing identification results for polynomial moments of Y.

[8]There is a misprint in Schennach (2007a), where it should be stated that $\gamma_o(\zeta) \neq 0$ instead of $\gamma(\zeta) \neq 0$. That paper relaxes the almost everywhere nonvanishing $\gamma_o(\zeta)$ assumption to $\gamma_o(\zeta) \neq 0$ almost everywhere in $[-\bar{\zeta}, \bar{\zeta}]$ for some $\bar{\zeta} \in [0, \infty]$ and $\gamma(\zeta) = 0$ for all $|\zeta| > \bar{\zeta}$.

[9]For instance, the Fourier transform of the cumulative distribution function of a standard normal is $\gamma(\zeta) = \pi\delta(\zeta) - \exp(-\zeta^2/2)/(\mathbf{i}\zeta)$, whose ordinary component is $\gamma_o(\zeta) = -\exp(-\zeta^2/2)/(\mathbf{i}\zeta)$, with $\delta(\zeta)$ denoting Dirac's delta function. More generally, the ordinary part can be defined by noting that any tempered distribution on $\mathbb{R}$ is the (generalized) derivative of some finite order of a continuous function. Wherever this derivative takes on a numerical value, it defines the ordinary part of the generalized function (as shown, e.g., in Laczkovich 1984, a numerical value for this derivative can be defined on a set that is sufficiently large to enable the determination of the ordinary part of the generalized function almost everywhere, which is sufficient for our purpose); elsewhere, the ordinary part can be conventionally set to zero. A few remarks may be of interest to some. The ordinary part $\gamma_o(\zeta)$ may not be a tempered distribution, but this does not invalidate the approach: $\gamma_o(\zeta)$ is a function in the usual sense, so standard operations on functions still hold. The singular part $\gamma_s(\zeta) \equiv \gamma(\zeta) - \gamma_o(\zeta)$ may not, in general, be a tempered distribution either, but is never used in isolation in the proof of identification.

[10]The multiplication of a generalized function, such as $\varepsilon_Y(\zeta)$, by an ordinary function, such as $1/\phi(\zeta)$, can be defined as in, e.g., Temple (1963, section 6). The assumption that both $|E[Y|Z = z]|$ and $|g(x^*)|$ are bounded by polynomials implies that $\varepsilon_Y(\zeta) = \gamma(\zeta)\phi(\zeta)$ and $\gamma(\zeta) = \varepsilon_Y(\zeta)/\phi(\zeta)$ are each associated with a unique tempered distribution, whose inverse Fourier transform always exists. It follows that the division by $\phi(\zeta)$ cannot cause $\varepsilon_Y(\zeta)/\phi(\zeta)$ to diverge in such a way that its inverse Fourier transform would fail to exist.

Schennach (2008) extends the instrumental variable approach to quantile regressions by working with indicator functions $\mathbf{1}(\cdot)$, using $E[\mathbf{1}(Y \leq y)|Z = z]$ and $E[X\,\mathbf{1}(Y \leq y)|Z = z]$ instead of $E[Y|Z = z]$ and $E[XY|Z = z]$ to construct analogs of Equations 44 and 45. This approach offers considerable advantages. First, the identification result delivers the entire joint distribution of Y and X^*. Second, by exploiting the boundedness of indicator functions (in combination with other techniques), the identification result can be phrased entirely in terms of regular (rather than generalized) functions, so that the identification result can be directly used to deliver a nonparametric estimator. This estimator is shown to be uniformly consistent.

4.4. Panel Data

Panel data offer unique opportunities for the identification of measurement error models because past or future observations can play the role of instruments or repeated measurements, so that no additional variable is needed to gain identification. This idea has been thoroughly explored in linear models (Griliches & Hausman 1986), but relatively little in nonlinear models because the problem is not as straightforward. Hausman et al. (1991) and Schennach (2004a) exploit the fact that future values of a mismeasured regressor can play the role of a repeated measurement. To see this, consider a simple panel model of the form

$$Y_t = g\left(X_t^*, \theta\right) + \Delta Y_t, \tag{53}$$

$$X_t = X_t^* + \Delta X_t, \tag{54}$$

$$X_{t+1}^* = X_t^* + U_{t+1}, \tag{55}$$

for $t = 1, \ldots, T$, where X_1^*, $\Delta X_1, \ldots, \Delta X_T, U_1, \ldots, U_T$, $\Delta Y_1, \ldots,$ and ΔY_T are mutually independent, $E[\Delta X_t] = 0$ and $E[\Delta Y_t] = 0$. It is clear that one can then use X_t and X_{t+1} as repeated measurements for X_t^*:

$$X_t = X_t^* + \Delta X_t, \tag{56}$$

$$X_{t+1} = X_{t+1}^* + \Delta X_{t+1} = X_t^* + (U_{t+1} + \Delta X_{t+1}), \tag{57}$$

where X_t^*, ΔX_t, and $(U_{t+1} + \Delta X_{t+1})$ are mutually independent, thus enabling the use of any Kotlarksi-type identity. This can be extended by replacing Equation 55 by a more general autoregressive process, e.g., $X_{t+1}^* = \rho X_t^* + U_{t+1}$, and first identifying ρ from $\text{Covar}[X_{t+2}, X_t]/\text{Covar}[X_{t+1}, X_t]$ before using X_t and X_{t+1}/ρ as a repeated measurement for X_t^*. However, it appears difficult to generalize the generating process for X_t^* much further without affecting the linear measurement structure.

An instrumental variable counterpart of this result is also difficult because X_t^* cannot be written as the sum of future or past observations plus an independent error term. For instance, for X_{t+1} or X_{t-1}, one has

$$X_t^* = X_{t+1} - (U_{t+1} + \Delta X_{t+1}), \tag{58}$$

$$X_t^* = X_{t-1} + (U_t - \Delta X_{t-1}), \tag{59}$$

in which none of the error terms in parentheses are independent from X_{t+1} or X_{t-1}. So the instrumental variable structure used in Hausman et al. (1991, 1995), Newey (2001), Wang & Hsiao (2011), and Schennach (2007a, 2008) is not applicable. However, the more general approach of Hu & Schennach (2008), discussed in Section 6.2, is applicable (for general generating processes

for X_t^*), as noted by Wilhelm (2010), because it does not rely on being able to write an instrument equation (of the form of Equation 38) with an independent disturbance. Furthermore, in this framework, individual-specific effects can be handled via first differences, considering $g(X_t^*, \theta) - g(X_{t-1}^*, \theta)$ as the regression function to identify with regressor vector (X_t^*, X_{t-1}^*).

Cunha et al. (2010) and Hu & Shiu (2013) identify a nonlinear dynamic panel data model (in which the fixed effect does not enter additively) by recognizing a connection with nonclassical error models to exploit the operator techniques from Hu & Schennach (2008). A rather different approach is devised by Wilhelm (2010), who provides a nearly closed-form solution to the identification problem for general generating processes for X_t^* by building on a clever combination of operator inversion methods (Newey & Powell 2003, Hall & Horowitz 2005, Carrasco et al. 2007, Darolles et al. 2011) with Fourier methods in quantile models (Schennach 2008).

5. METHODS WITHOUT AUXILIARY VARIABLES

In some applications, suitable auxiliary variables, such as repeated measurements or instruments, are not available, and it is of interest to investigate what, if anything, can be learned solely from the data on the mismeasured regressor X and the dependent variable Y in a regression setting. Two possible approaches are surveyed here: one in which stronger independence assumptions are made to secure identification and one in which weak assumptions are maintained at the expense of possibly obtaining only bounds on the coefficients of interest instead of point estimates.

5.1. Higher-Order Information

If it is plausible to assume that X^*, ΔX, and ΔY are mutually independent, then it is well-known that the standard regression model in Equation 3 with a linear specification is identified under simple conditions. Mutual independence implies a number of moment conditions that provide a large number of nonlinear equations that can be solved to secure identification. More generally, the key idea is to use the full distribution of the variables, rather than only their means and covariances, to gain identification. Since the seminal works of Geary (1942) and Reiersol (1950), many authors have exploited this idea to identify and estimate a linear specification (e.g., Kendall & Stuart 1979; Pal 1980; Kapteyn & Wansbeek 1983; Cragg 1997; Dagenais & Dagenais 1997; Lewbel 1997; Erickson & Whited 2000, 2002; Bonhomme & Robin 2009; Lewbel 2012; and the many references therein).

The basic idea is to consider higher-order moments of X and Y and note that independence implies that these moments can be expressed in terms of products of moments of X^* and moments of the errors ΔX and ΔY. These expressions can be solved for the slope parameter θ in terms of observable moments. One simple result of this kind is that if $E[(X^* - E[X^*])^3] \neq 0$, then

$$\theta = \frac{\text{Covar}[Z, Y]}{\text{Covar}[Z, X]}, \tag{60}$$

where $Z = (X - E[X])(Y - E[Y])$. This expression has a natural instrumental variable interpretation, in which the instrument is constructed from the variables themselves and does not need to be externally provided. It is possible to relax the assumption that $E[(X^* - E[X^*])^3] \neq 0$ by using fourth-order mixed moments, but then the restriction $E[(X^* - E[X^*])^4] \neq 3E[(X^* - E[X^*])^2]$ is required. Going to even higher moments keeps slightly expanding the set of allowed distributions, but some unidentified cases always remain, as formally shown by Reiersol (1950) using the CF.

When working with higher-order moments of sums of independent variables, it is considerably more convenient to instead work with the related concept of cumulants (Geary 1942), defined as

derivatives of the logarithm of the CF, as the cumulants of a sum of independent random variables are simply the sum of their corresponding cumulants. Bonhomme & Robin (2009) considerably generalize the treatment of Geary (1942) to multivariate factor models (introduced in Equation 37), enabling identification of the factor loadings with fewer measurements than from covariance information alone (Anderson & Rubin 1956), as discussed in Section 4.2.

Although linear specifications have historically received considerable attention, the question of nonparametric identification of the regression model in Equation 3 has only recently been answered (Schennach & Hu 2013). The answer turns out to be remarkably simple: Among all (sufficiently regular) specifications, only a small parametric family is not identified. To state this result, let us assume that (*a*) the marginal CFs of ΔX, ΔY, X^*, and $g(X^*)$ do not vanish anywhere, (*b*) $f_{X^*}(x^*)$ exists, (*c*) $g(x^*)$ is continuously differentiable, and (*d*) $g'(x^*)$ vanishes (at most) at a finite number of points, and when that happens $f_{X^*}(x^*)$ is continuous and nonvanishing. There are four mutually exclusive cases:

1. If $g(x^*)$ is not of the form
$$g(x^*) = a + b \ln\left(e^{cx^*} + d\right) \tag{61}$$
for some constants $a, b, c, d \in \mathbb{R}$, then the model is identified.
2. If $g(x^*)$ is of the form of Equation 61 with $d > 0$, then the model is not identified if and only if X^* has a density of the form
$$f_{X^*}(x^*) = A \exp\left(-Be^{Cx^*} + CDx^*\right)\left(e^{Cx^*} + E\right)^{-F} \tag{62}$$
with $C \in \mathbb{R}$, $A, B, D, E, F \in [0, \infty)$, and ΔY has a type I extreme value factor[11] [i.e., a density of the general form $f(u) = K_1 \exp(K_2 \exp(K_3 u) + K_4 u)$].
3. If $g(x^*)$ is of the form of Equation 61 with $d < 0$, then case 2 above applies, after permuting the roles of X and Y.
4. If $g(x^*)$ is linear (i.e., of the form of Equation 61 with $d = 0$), then the model is not identified if and only if X^* is normally distributed and either ΔX or ΔY has a normal factor (this is the case covered in Reiersol 1950).

What is remarkable is that, among all possible (sufficiently regular) specifications, the nonidentified family is parametric with only four parameters (a, b, c, and d in Equation 61). Within that family, whenever $d \neq 0$, the model is actually locally identified (there are only two observationally equivalent models in this case: one with $d > 0$ and one with $d < 0$). Additionally, the notoriously difficult case of normal errors is in fact identified for virtually all specifications, with the linear case ($d = 0$) being the only exception.

The idea of using the full distribution of the variables to gain point identification without relying on auxiliary variables has also been used in regression models with discrete mismeasured regressors (Chen et al. 2009). In that case, thanks to the discrete nature of the problem, the identification can be reduced to a finite, albeit complex, system of equations.

5.2. Bounding

What if auxiliary variables are not available and independence cannot plausibly be assumed? Then, unfortunately, one may have to be willing to abandon the hopes of identifying a single value of the parameter and instead settle for a set of possible values. The concept of set identification

[11] A random variable is said to have a distribution F as a factor if it can be written as the sum of two independent random variables (which may be degenerate), one of which has distribution F.

has some of its roots in the early literature on measurement error (Frisch 1934), and recently, set identification has gathered considerable attention in econometrics (e.g., Manski 1990, 2003; Manski & Tamer 2002; Chernozhukov et al. 2007a).

The bounds on the slope coefficient θ obtained by Frisch (1934) for the linear regression model are based on the idea that the variance of a variable must be nonnegative. Under the assumption of mutual uncorrelatedness of X^*, ΔX, and ΔY, the full covariance structure of X and Y can be expressed in terms of the variances of the unobservable variables $\sigma^2_{X^*}$, $\sigma^2_{\Delta X}$, and $\sigma^2_{\Delta Y}$:

$$\text{Var}[Y] = \theta^2\sigma^2_{X^*} + \sigma^2_{\Delta Y}, \tag{63}$$

$$\text{Var}[X] = \sigma^2_{X^*} + \sigma^2_{\Delta X}, \tag{64}$$

$$\text{Covar}[X, Y] = \theta\sigma^2_{X^*}. \tag{65}$$

Combining Equations 64 and 65 and using the constraint $\sigma^2_{\Delta X} \geq 0$ yield (if $\theta \geq 0$)

$$\theta \geq \frac{\text{Covar}[X, Y]}{\text{Var}[X]}, \tag{66}$$

whereas combining Equations 63 and 65 and using the constraint $\sigma^2_{\Delta Y} \geq 0$ yield (if $\theta \geq 0$)

$$\theta \leq \frac{\text{Var}[Y]}{\text{Covar}[X, Y]}. \tag{67}$$

The directions on the inequalities in Equations 66 and 67 are reversed if $\theta < 0$. These bounds have the straightforward interpretation as the slope coefficient of the forward regression of Y on X (Equation 66) and of the reciprocal of the slope coefficient of the reverse regression of X on Y (Equation 67). These bounds are sharp under the uncorrelatedness of X^*, ΔX, and ΔY.

This result has been elegantly generalized by Klepper & Leamer (1984) to multivariate linear regression in which all regressors $(X^*_1, \ldots, X^*_{d_X})$ are measured with error. It suffices to sequentially use each of the $d_{X^*}+1$ observed variables $(Y, X_1, \ldots, X_{d_X})$ as the dependent variable and remaining variables as regressors. After rearrangement, each regression line $k = 1, \ldots, d_{X^*}$ can be cast in the form $Y^* = \sum_{j=1}^{d_{X^*}} \theta_j^{(k)} X^*_j$ (where $Y^* \equiv Y - \Delta Y$), so that the regression coefficient vectors $\theta^{(k)}$ all have the same units. Then, the set obtained by taking the convex hull of $\theta^{(1)}, \ldots, \theta^{(d_{X^*})}$ is a consistent estimator of the identified set, the true set of possible values of the regression coefficients, provided the $\theta^{(k)}$ all lie in the same orthant.[12] Otherwise, the set is unbounded along some direction(s). The issue of potentially unbounded identified sets can be mitigated by a priori plausible restrictions on the variance and the correlation structure of the measurement error (Klepper & Leamer 1984, Erickson 1993).

Unfortunately, the derivation of sharp bounds in the general nonlinear measurement error problem is usually analytically intractable due to potentially nonmonotone relationships between the observed and unobserved variables. For this reason, it is more useful to use general simulation-based methods to derive the bounds numerically. Schennach (2014) proposes such a method for a general class of latent variable models that nests measurement error problems as a special case. Other recent methods aimed at handling related latent variable models include Galichon & Henry (2013), Ekeland et al. (2010), and Beresteanu et al. (2011).

Schennach (2014) seeks to find the value(s) of a parameter vector θ that satisfies a set of moment conditions that are known to hold in the population:

$$E[m(U, Z, \theta)] = 0, \tag{68}$$

[12] An orthant is a set of vectors whose elements share the same pattern of signs (it is the multivariate generalization of a quadrant).

where m is a d_m-dimensional vector of nonlinear measurable functions depending on the parameter vector θ, on an unobserved random vector U, and on an observed random vector Z. Naturally, it is the unobservable U that enables the treatment of measurement error models (typically, one can then set $U = X^*$ and express the errors ΔX, etc., as a function of X^* and the observable variables). The method is called entropic latent variance integration via simulation (ELVIS) and avoids any parametric assumptions (beyond the given functional form of m), without introducing any infinite-dimensional nuisance parameters, through the use of a low-dimensional representation of the identification problem.

To describe this method, let π denote the distribution of the observables Z and μ denote a conditional distribution of the unobservables U given $Z = z$, whereas their joint distribution is denoted by $\mu \times \pi$. Let $\mathcal{U}$ denote the support of U conditional on $Z = z$ [where it may be assumed without loss of generality that $\mathcal{U}$ does not depend on z, as this dependence could be incorporated into $m(u, z, \theta)$]. Expectations are calculated under the distribution specified as a subscript. The key identity is as follows: For any θ and any distribution π,

$$\inf_{\mu} \left\| E_{\mu \times \pi} \left[m\left(U, Z, \theta\right) \right] \right\| = 0 \tag{69}$$

if and only if

$$\inf_{\gamma \in \mathbb{R}^{d_m}} \left\| E_{\pi} \left[\tilde{m}\left(Z, \theta, \gamma\right) \right] \right\| = 0, \tag{70}$$

where

$$\tilde{m}\left(z, \theta, \gamma\right) \equiv \frac{\int m\left(u, z, \theta\right) \exp\left(\gamma' m\left(u, z, \theta\right)\right) \mathrm{d}\rho\left(u|z\right)}{\int \exp\left(\gamma' m\left(u, z, \theta\right)\right) \mathrm{d}\rho\left(u|z\right)}, \tag{71}$$

with ρ a user-specified conditional distribution such that (*a*) $\rho(\cdot|z)$ is supported on $\mathcal{U}$ at each z and (*b*) $E_{\pi}[\ln E_{\rho}[\exp(\gamma' m(U, Z, \theta))|Z]]$ exists and is twice differentiable in γ for all $\gamma \in \mathbb{R}^{d_m}$ and all θ. Measures $\rho(u|z)$ satisfying the above restrictions are easy to construct. For instance, if $\mathcal{U}$ is compact and sufficiently regular, $\rho(u|z)$ can simply be set to the Lebesgue measure on $\mathcal{U}$. More generally, ρ merely needs to have the right support and sufficiently thin tails. A general recipe for constructing a suitable ρ is given in Schennach (2014, proposition 1).

This identity is useful because the original problem of interest (Equation 69), which involves optimization over an infinite-dimensional quantity μ, has been replaced by a finite-dimensional optimization (Equation 70). This simplification is made possible by the fact that the so-called parametric least-favorable entropy-maximizing family of distributions [proportional to $\exp(\gamma' m(u, z, \theta))\mathrm{d}\rho(u|z)$] used to compute the expectation over U in Equation 71 is such that it spans exactly the same range of values of moments (as γ varies) as the original, fully nonparametric, problem (as the distribution of U varies). This is true for any distribution π (even for the empirical distribution of the sample) and for any choice of ρ (provided it satisfies the stated conditions). Consequently, the choice of ρ has no effect on the properties of any estimator based on the moment conditions $E[\tilde{m}(z, \theta, \gamma)] = 0$, even in finite samples (as the optimization over γ would yield the same profiled objective function in terms of θ).

The intuition behind this result is that Equation 69 has potentially an infinite number of solutions. However, because one only needs to find one, one can rank distributions according to some criterion (here, their entropy) and convert an existence problem into an optimization problem under the constraint in Equation 68. This constrained entropy maximization problem has a unique solution, which turns out to have the convenient form of Equation 71.

The simplest way to evaluate the integral in Equation 71 defining the moment function is to draw random vectors u_j, $j = 1, \ldots, R$, from a density proportional to $\exp(\gamma' m(u, z, \theta))\mathrm{d}\rho(u|z)$

using, e.g., the Metropolis algorithm and calculate the average

$$\frac{1}{R}\sum_{j=1}^{R} m\left(u_j, z, \theta\right) \tag{72}$$

for a sufficiently large R. A nice feature of the Metropolis algorithm is that it automatically takes care of the normalization integral in the denominator of Equation 71. This simulation-based approach essentially amounts to plugging in a parametric least-favorable entropy-maximizing family of distributions into the method of simulated moments (MSM) (McFadden 1989, Pakes & Pollard 1989), so ELVIS can be seen as a semiparametric generalization of the MSM.

Averaging over the unobservables then provides a conventional vector of moment conditions $E[\tilde{m}(Z, \theta, \gamma)] = 0$ involving only observable variables that is equivalent to the original moment condition in Equation 68. As a result, solving for the parameter θ of interest and for the nuisance parameter γ can be accomplished through a variety of standard techniques: conventional GMM (Hansen 1982) or any of its one-step alternatives, such as empirical likelihood (Owen 1988), generalized empirical likelihood (e.g., Owen 1990, Qin & Lawless 1994, Kitamura & Stutzer 1997, Imbens et al. 1998, Newey & Smith 2004), or exponentially tilted empirical likelihood (Schennach 2007b). Existing generic inference techniques for set-identified models (e.g., Chernozhukov et al. 2007a) can then be used. The ELVIS objective function bypasses the complex task of establishing point or set identification of the model by providing a vector of moment conditions that are, by construction, satisfied (asymptotically) over the identified set, whether it is a single point or a larger set.

The ELVIS method can be used in a measurement error context (see Schennach 2014, figure 6) to estimate the identified set in general nonlinear regression models, while assuming only very weak uncorrelatedness assumptions regarding the errors and without using any instruments or repeated measurements. Furthermore, ELVIS automatically adapts when a model transitions from being point identified to set identified, as the data-generating process changes. For instance, in a linear regression with measurement error estimated using higher-order moments, Schennach (2014, figure 5) shows that the objective function flattens over an interval (which is indicative of set identification[13]) when the distribution of the true regressor becomes normal, whereas it is strongly curved for the point-identified nonnormal case. ELVIS thus nests both the higher-order moment treatment and the bounding treatment of the standard linear errors-in-variables model. ELVIS is also a natural approach to handle a combination of measurement error with other latent variable problems (e.g., censoring, truncation, interval-valued data, limited dependent variables, panel data with nonseparable correlated individual-specific heterogeneity, and various game-theoretic models). Schennach (2014) also provides extensions to conditional mean and independence restrictions.

6. NONCLASSICAL MEASUREMENT ERROR

The classical measurement error assumptions have come under growing scrutiny in recent years (Bound et al. 2001, Bollinger 1998, Hyslop & Imbens 2001). While it is easy to point out the weaknesses of the classical assumptions, it is much more difficult to find solutions. As discussed in Section 3.2, validation data are, in principle, a general answer, but, in practice, rarely available. There is therefore considerable interest in developing practical methods to handle fairly general types of nonclassical measurement errors using, instead, more commonly available auxiliary information familiar to economists, such as instrumental variables or repeated measurements. A

[13] It is possible for a set-identified model to be associated with an objective function that is not perfectly flat over a set in finite samples—this happens when the so-called degeneracy property does not hold (Chernozhukov et al. 2007a).

special feature of nonclassical errors is that, unlike classical errors, they induce bias in a standard regression even when only the dependent variable is mismeasured. Hence, the methods described below are either applicable to both mismeasured regressors and mismeasured dependent variables or specifically target dependent variable measurement error.

6.1. Misclassification

When a discrete variable X^* is measured with error (and its measurement X is also discrete) this variable is said to be misclassified. Misclassification is considered nonclassical because, when the number of possible values X can take is finite, extreme values of X^* can only be mismeasured in one direction, so a zero-mean error (conditional on the true value X^*) is impossible.

Mahajan (2006) shows that a binary instruments can be used to identify and estimate an index model with a misclassified binary regressor and other perfectly measured regressors (which may be continuously distributed), whereas Lewbel (2007) considers treatment effect models when the treatment (a binary variable) is misclassified and an instrument is available. Hu (2008) goes beyond the binary case and solves the identification problem for general misclassified discrete regressors by exploiting an analogy between the identification problem and matrix diagonalization (this interesting approach is not described here, as it can be seen as a finite-dimensional version of the approach described in Section 6.2). A very similar technique is used by Bonhomme et al. (2016) to show identification of some mixture models.

In parametric discrete response models with misclassification in the response variable, Cameron et al. (2004), Hausman et al. (1998), and Li et al. (2003) observe that identification without additional information is made possible by the following fact: The errors in the response produce distortions in the observable distribution of the mismeasured response that are distinct from the distributions that could be generated from the underlying parametric model of the true unobserved response. This can be exploited within a maximum likelihood procedure with flexible functional forms while employing simulations when integration cannot be performed analytically (Li et al. 2003, Cameron et al. 2004). Hausman et al. (1998) focus on the binary response case and consider a semiparametric treatment (thus avoiding a parametric specification of the measurement error distribution) using a combination of a maximum rank correlation estimator and isotonic regression techniques.

The problem of identifying the distribution of an unobserved discrete variable from misclassified repeated measurements has received considerable attention in the statistics literature (see, e.g., Kruskal 1989 and Kolda & Bader 2009 for reviews). One of the most significant result is Kruskal's (1977) tensor array decomposition, which is described here. Assume that one has access to three potentially misclassified measurements (X, Y, and Z) of X^* satisfying a conditional independence assumption: Conditional on X^* taking a specific value, the variables X, Y, and Z are independent. Expressed in terms of probability mass functions (denoted by f with appropriate subscripts), this condition reads $f_{X,Y,Z|X^*}(x, y, z|x^*) = f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*)$, which enables us to easily express the observable quantity $f_{X,Y,Z}(x, y, z)$ in terms of the unobservable quantities of interest, $f_{X|X^*}(x|x^*)$, $f_{Y|X^*}(y|x^*)$, $f_{Z|X^*}(z|x^*)$, and $f_{X^*}(x^*)$. In the case of discrete variables, one has

$$f_{X,Y,Z}(x, y, z) = \sum_{x^* \in \mathcal{X}} f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*), \tag{73}$$

where $\mathcal{X}$ denotes the discrete support of the distribution of X^*. One can associate each quantity on the right-hand side with (possibly rectangular) matrices, $A_{ir} \equiv f_{X|X^*}(i|r)$, $B_{jr} \equiv f_{Y|X^*}(j|r)$, and $C_{kr} \equiv f_{Z|X^*}(k|r) f_{X^*}(r) = f_{Z,X^*}(k, r)$, and relabel the support points of each variables as consecutive

integral numbers, without loss of generality. Similarly the left-hand side can be associated with a three-way array $T_{ijk} \equiv f_{X,Y,Z}(i, j, k)$.

The following definition is useful: The Kruskal rank of a matrix M, denoted K_M, is the largest k such that any k columns of M are not colinear. Note how this definition differs from the usual rank, which is the largest k such that there merely exists one choice of k columns that are noncolinear.

Kruskal's result can now be stated: If a three-way array T admits the decomposition

$$T_{ijk} = \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr} \tag{74}$$

with

$$K_A + K_B + K_C \geq 2R + 2, \tag{75}$$

then any other triple of matrices $(\tilde{A}, \tilde{B}, \tilde{C})$ such that

$$T_{ijk} = \sum_{r=1}^{R} \tilde{A}_{ir} \tilde{B}_{jr} \tilde{C}_{kr} \tag{76}$$

will satisfy $\tilde{A} = AD_A P$, $\tilde{B} = BD_B P$, and $\tilde{C} = CD_C P$, where D_A, D_B, and D_C are diagonal matrices satisfying $D_A D_B D_C = I$, and P is a permutation matrix. When the matrices A, B, and C represent probabilities, one can use the fact that $\sum_i A_{ir} = \sum_x f_{X|X^*}(x|x^*) = 1$ to uniquely determine D_A. Similarly, $\sum_j B_{jr} = 1$ uniquely determines D_B. Next, D_C is recovered from $D_C = D_A^{-1} D_B^{-1}$ (because $\sum_k C_{kr} \neq 1$ in general, as C_{kr} represents a joint rather than a marginal distribution).

The result effectively states that, under the rank condition given in Equation 75, the observable $f_{X,Y,Z}(x, y, z)$ can be uniquely associated with one tuple of unobserved probability mass functions $f_{X|X^*}(x|x^*)$, $f_{Y|X^*}(y|x^*)$, and $f_{Z,X^*}(z, x^*)$, apart from a trivial reordering of the unobservable x^* (via the permutation matrix P). This result is particularly adapted to factor models because all variables of the model are treated in a symmetric fashion. The Kruskal-based approach to the identification of discrete models is receiving increasing attention in the statistics and econometrics literature (e.g., Allman et al. 2009, Bonhomme et al. 2016). Note that Kruskal's condition is not necessary—slightly weaker conditions have been obtained (Stegeman 2009) at the expense of losing a symmetric treatment of the three matrices A, B, and C.

Identification results for discrete variables do not imply identification of the corresponding model with continuous variables via a simple limiting argument. To illustrate this, let $\mathrm{Id}(M)$ denote a function equal to 1 if model M is identified and zero otherwise. This function is necessarily a discontinuous function of M because its range is discrete. Consequently, for a sequence of models M_n, one does not have $\lim_{n\to\infty} \mathrm{Id}(M_n) = \mathrm{Id}(\lim_{n\to\infty} M_n)$ in general. Therefore, when constructing a sequence of identified discrete models converging to a continuous model, one cannot conclude that the limiting continuous model is identified. A proper way to approach the continuous case would be to instead derive an identified set for the discrete case and show that this set converges to a point (e.g., in the Hausdorff metric) as the discretization becomes finer.

6.2. Continuously Distributed Variables

Hu & Schennach (2008) establish that general nonlinear models with continuous variables contaminated by nonclassical measurement errors can be identified via an auxiliary variable Z satisfying assumptions sufficiently general to cover both repeated measurements and instruments. (Additional, perfectly measured regressors can easily be included in this framework by conditioning on them.) They show that, under suitable conditions outlined below, for a given true observed

density $f_{YX|Z}$, the equation

$$f_{YX|Z}(y, x|z) = \int f_{Y|X^*}(y|x^*) f_{X|X^*}(x|x^*) f_{X^*|Z}(x^*|z) \, \mathrm{d}x^* \tag{77}$$

admits a unique solution $(f_{Y|X^*}, f_{X|X^*}, f_{X^*|Z})$. A similar result holds for

$$f_{YXZ}(y, x, z) = \int f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*) \, \mathrm{d}x^*. \tag{78}$$

This general phrasing of the identification result implies that any model that would be identified from $f_{Y|X^*}(y|x^*)$ or $f_{YX^*}(y, x^*)$ (if X^* were observed) is identified from the knowledge of the observed densities $f_{YX|Z}(y, x|z)$ or $f_{YXZ}(y, x, z)$. To fix the ideas, one can think of Y as the dependent variable and X^* as the regressor, but this assignment is not the only possible choice. For instance, all variables (X, Y, Z) can also be considered on a more symmetric footing to form a general nonlinear (dynamic) factor model: Cunha et al. (2010) explain how the distribution of X^* and the general nonlinear and nonseparable relationships between the factor vector X^* and (X, Y, Z) can be recovered from the identified densities $f_{Y|X^*}(y|x^*)$, $f_{X|X^*}(x|x^*)$, $f_{Z|X^*}(z|x^*)$, and $f_{X^*}(x^*)$, using widely used normalizations borrowed from the literature on nonseparable models (e.g., Matzkin 2003), which is discussed further in the next section.

Handling nonclassical errors in continuous variables requires a considerably more advanced technical apparatus than in the discrete case because the unknowns are infinite dimensional (they are whole functions, not just vectors). One important notion is the concept of a linear operator (see, e.g., Carrasco et al. 2007), which can be seen as the generalization of a matrix to infinite dimensions. To outline the identification result, let us first define, for any conditional density $f_{V|U}(v|u)$, an operator $L_{V|U}$ mapping a (sufficiently regular) arbitrary function q to the function:

$$\left[L_{V|U} q\right](v) \equiv \int f_{V|U}(v|u) q(u) \, \mathrm{d}u. \tag{79}$$

The conditions needed for identification are that (*a*) the random vectors X, Y, and Z are mutually independent conditional on X^*, (*b*) the operators $L_{X|X^*}$ and $L_{Z|X^*}$ are injective,[14] (*c*) the distributions of Y given $X^* = x_1^*$ and of Y given $X^* = x_2^*$ differ whenever $x_1^* \neq x_2^*$, and (*d*) there exists a known functional M such that $M[f_{X|X^*}(\cdot|x^*)] = x^*$. Assumption (*a*) is a fairly natural exclusion restriction. Assumption (*b*) is a generalization of familiar rank conditions for matrices to operators. It is similar to assumptions commonly made (sometimes under the name of completeness or nonsingularity) in the literature on nonparametric instrumental variable methods (Newey & Powell 2003, Chernozhukov & Hansen 2005, Hall & Horowitz 2005, Carrasco et al. 2007, Darolles et al. 2011). In the special case of convolution operators, simple sufficient conditions for injectivity have been obtained (Mattner 1993, Hu & Ridder 2010, d'Haultfoeuille 2011), and progress is being made for more general operators (e.g., Andrews 2011, Hu et al. 2015). Whereas the injectivity assumption (*b*) requires that the dimensions of X^*, X, and Z be the same, assumption (*c*) is weaker than injectivity and can be satisfied even if Y is scalar.

Remarkably, assumptions (*b*) and (*c*) jointly demand that $d_X + d_Y + d_Z \geq 2d_{X^*} + 1$, which is the same dimensionality constraint as in a linear factor model (Anderson & Rubin 1956). Assumption (*d*) generalizes classical measurement error (corresponding to M being the mean) to other types of centering restrictions (e.g., the mode, median, or any other quantile). As discussed by Hu & Schennach (2008), median and mode restrictions are well supported by existing validation data evidence (Bound & Krueger 1991). In addition, such assumptions are robust to other data

[14]Or, equivalently, $L_{X|X^*}$ and $L_{Z|X}$ are injective.

problems: For instance, under weak conditions, the mode is unaffected by truncation, whereas the median is unaffected by censoring.

The proof of identification in Hu & Schennach (2008) can be outlined as follows. Assumption (*a*) directly implies the integral Equation 77. This equation can be cast as an operator equivalence relationship:

$$L_{y;X|Z} = L_{X|X^*} D_{y;X^*} L_{X^*|Z}, \tag{80}$$

where $L_{y;X|Z}$ is defined analogously to $L_{X|Z}$, with $f_{X|Z}$ replaced by $f_{Y,X|Z}(y,\cdot|\cdot)$ for a given y, and $D_{y;X^*}$ is the diagonal operator mapping a function $q(x^*)$ to the function $f_{Y|X^*}(y|x^*)q(x^*)$, for a given y. Next, note that the equivalence $L_{X|Z} = L_{X|X^*}L_{X^*|Z}$ also holds [as $f_{X|Z}(x|z) = \int f_{X|X^*}(x|x^*)f_{X^*|Z}(x^*|z)\mathrm{d}x^*$, again by conditional independence]. Isolating $L_{X^*|Z}$ to yield

$$L_{X^*|Z} = L^{-1}_{X|X^*} L_{X|Z}, \tag{81}$$

substituting it into Equation 80, and rearranging, one obtains

$$L_{y;X|Z} L^{-1}_{X|Z} = L_{X|X^*} D_{y;X^*} L^{-1}_{X|X^*}, \tag{82}$$

where all inverses can be shown to exist over suitable domains[15] under the injectivity assumptions made.

Equation 82 states that the operator $L_{Y;X|Z}L^{-1}_{X|Z}$ admits a spectral decomposition (eigenvalue-eigenfunction decomposition). The operator to be diagonalized is defined in terms of observable densities, whereas the resulting eigenvalues $f_{Y|X^*}(y|x^*)$ and eigenfunctions $f_{X|X^*}(\cdot|x^*)$ (both indexed by x^*) provide the unobserved densities of interest. To ensure uniqueness of this decomposition, Hu & Schennach (2008) employ four techniques. First, a powerful result from spectral analysis (Dunford & Schwartz 1971, theorem XV 4.5) ensures uniqueness up to some normalizations. Second, the a priori arbitrary scale of the eigenfunctions is fixed by the requirement that densities must integrate to one. Third, to avoid any ambiguity in the definition of the eigenfunctions when degenerate eigenvalues are present, one can use assumption (*c*) and the fact that the eigenfunctions [which do not depend on y, unlike the eigenvalues $f_{Y|X^*}(y|x^*)$] must be consistent across different values of the dependent variable y. Finally, to uniquely determine the ordering and indexing of the eigenvalues and eigenfunctions, one can invoke assumption (*d*): Consider another variable $\tilde{x}^*$ related to x^* through $x^* = R(\tilde{x}^*)$, and note that

$$M\left[f_{X|\tilde{X}^*}(\cdot|\tilde{x}^*)\right] = M\left[f_{X|X^*}(\cdot|R(\tilde{x}^*))\right] = R(\tilde{x}^*), \tag{83}$$

which is only equal to $\tilde{x}^*$ if R is the identity function. Observe that in discrete models, the centering restrictions can be considerably weakened to $M[f_{X|X^*}(\cdot|x^*)]$ strictly increasing in x^* because, in finite dimension, ordering the eigenvectors is sufficient, whereas in infinite dimensions, it is possible to reparameterize the eigenfunctions without changing their order [e.g., $x^* = (\tilde{x}^*)^3$].

The four above steps ensure that the diagonalization operation uniquely specifies the unobserved densities $f_{Y|X^*}(y|x^*)$ and $f_{X|X^*}(x|x^*)$ of interest. Next, Equation 81 implies that $f_{X^*|Z}(x^*|z)$ is also identified. Because the identities in Equations 82 and 81 use and provide the same information as Equation 77, this establishes uniqueness of the solution to Equation 77. Equation 78 follows by similar manipulations. Hu & Schennach (2008) also suggest an estimator based on Equation 77 or 78 obtained by substituting series approximations for the unknown densities.

Interestingly, this identification result reaches a continuous analog of Kruskal's rank bounds for discrete variables. To see this, first rewrite Kruskal's bound (Equation 75) as

$$K_B \geq (R - K_A) + (R - K_C) + 2. \tag{84}$$

[15] The injectivity of $L_{Z|X}$ is related to the existence of $L^{-1}_{X|Z}$ (Hu & Schennach 2008, lemma 1).

In Hu & Schennach (2008), the operators associated with two of the conditional densities ($f_{X|X^*}$ and $f_{Z|X^*}$) are injective, so one could consider them to have full rank. In Kruskal's inequality (Equation 75), one thus (informally) has $R - K_A = 0$ and $R - K_C = 0$. For the third variable Y, one only needs any two columns $f_{Y|X^*}(\cdot|x_1^*)$ and $f_{Y|X^*}(\cdot|x_2^*)$ to be different, which secures a Kruskal rank of two (as any two distinct columns are necessarily linearly independent), and hence $K_B = 2$. One then has $2 \geq 0 + 0 + 2$, thus reaching Kruskal's bound. Of course, the above argument is heuristic as the subtraction of infinities is not handled very formally (e.g., $R - K_A = 0$). Nevertheless, this observation suggests that the conditions in Hu & Schennach (2008) cannot be made significantly weaker. However, their result does not possess the symmetry in the indicator variables (X, Y, Z) that Kruskal's result has. The difficulty lies in even defining the Kruskal rank when the number of dimensions is infinite in the general case where the operator is neither injective nor has a Kruskal rank of two.

Operator diagonalization techniques generalize Fourier transforms, which only diagonalize convolution operators (whose eigenfunctions are always known as they are just complex exponentials). Thanks to this extra generality, operator diagonalization techniques are now being used to solve an increasing number of identification problems with nonclassical disturbances, such as dynamic models (Hu & Shum 2012, 2013), dynamic factor models (Cunha et al. 2010), two-sample combination methods (Carroll et al. 2010), and nonlinear panel data models via quantile restrictions (Arellano & Bonhomme 2015), with heterogeneity and selection (Sasaki 2015) and with interactive fixed effects (Freyberger 2012), as well as Berkson-type errors (Schennach 2013b), discussed further below.

Operator diagonalization is not the only approach, however. Hu & Sasaki (2015) adapt Kotlarski's approach (discussed in Section 4.1) to obtain a closed-form solution to a measurement system of the form

$$Y = g(X^*) + \Delta Y, \tag{85}$$

$$X = X^* + \Delta X, \tag{86}$$

$$Z = \sum_{p=0}^{P} \gamma_p (X^*)^p + \Delta Z, \tag{87}$$

where $g(\cdot)$ is a generic unknown nonlinear function, γ_p are unknown polynomial coefficients, $E[\Delta X] = 0$, $E[\Delta Z] = 0$, $E[\Delta Y | X^*, \Delta X, \Delta Z] = 0$, and X^*, ΔX, and ΔZ are mutually independent. Despite the independence assumptions, two of the measurements (Y and Z) are nonclassical in that they are nonlinearly related to X^*. Their approach proceeds in four steps. First, expectations of products of Y, X, and Z^q are used to identify the γ_p and moments of ΔX. Next, the latter are used to recover $E[(X^*)^q \exp(i\zeta Z^*)]$ [with $Z^* \equiv \sum_{p=0}^{P} \gamma_p (X^*)^p$] from the observed $E[(X)^q \exp(i\zeta Z)]$. Then, one forms the linear combination $\sum_{p=0}^{P} \gamma_p E[(X^*)^q \exp(i\zeta Z^*)]$ to obtain $E[Z^* \exp(i\zeta Z^*)]$, which allows the use of Kotlarski-type techniques. Finally, the distribution of X^* is inferred from the one of Z^*, exploiting a monotonicity assumption and the knowledge of the (now identified) γ_p.

6.3. Nonseparable Error Models

The idea that disturbances need not enter economic models in an additive fashion has gained considerable acceptance over the past decade. Nonseparability has close ties to the concepts of average effects, treatment effects, and quantiles (see, e.g., Chesher 2003; Matzkin 2003, 2008; Heckman & Vytlacil 2005; Chernozhukov et al. 2007b).

The basic setup is a model of the form

$$Y = g(X, \Delta Y), \tag{88}$$

where X and Y are observed. The fundamental issue is that, of course, changes in the distribution $F_{\Delta Y}$ of the unobservable ΔY can be offset by corresponding changes in the general nonlinear function $g(\cdot, \cdot)$. To see this, consider a one-dimensional ΔY and a strictly monotone function t. Then, the model

$$Y = \tilde{g}\left(X, \Delta \tilde{Y}\right), \tag{89}$$

where $\tilde{g}(x, u) = g(x, t(u))$ and $\Delta \tilde{Y} = t^{-1}(\Delta Y)$, is observationally equivalent to Equation 88 because $Y = \tilde{g}(X, \Delta \tilde{Y}) = g(X, t(t^{-1}(\Delta Y))) = g(X, \Delta Y)$. The two common solutions to this problems are (*a*) the focus on average effects that are independent of the particular parameterization of the model (e.g., Chesher 2003) or (*b*) the use of normalizations [of either $g(\cdot, \cdot)$ or $F_{\Delta Y}$] (e.g., Matzkin 2003). Handling nonseparable disturbances acquires an even more challenging nature when the disturbance ΔY is endogenous and an instrument must be used to secure identification. There is strong interest in developing methods to handle such models (e.g., Abadie et al. 2002; Chesher 2003; Matzkin 2003, 2008; Chernozhukov & Hansen 2005; Heckman & Vytlacil 2005; Chernozhukov et al. 2007b; Hoderlein & Mammen 2007).

These efforts, however, assume that all variables (except the disturbance) are observed. Extensions to unobserved variables can be accomplished in two steps. First, the joint distribution of all observable variables (typically repeated measurements or indicators within a factor model) is used to determine the distribution of the unobservables with suitable measurement error techniques. Second, nonseparable error techniques can be employed to determine the nonseparable relationships existing between these unobserved variables by using the (now identified) distribution of the unobservables as an input, as if it were observed. In an example of this approach, Cunha et al. (2010) use the identification result of Hu & Schennach (2008) (which allows for nonclassical and therefore nonseparable disturbances) to uncover the distribution of the unobservables and use the nonseparable techniques of Matzkin (2003) to identify the functions that relate these variables to each other.

Other approaches combine these two conceptual steps in different ways. Matzkin (2003) considers the identification of a wide range of nonseparable models of survey response errors, exploiting either validation data or repeated measurements. The identification and estimation of various nonseparable models and associated average effects from repeated measurements have been investigated by Schennach et al. (2012) and Song et al. (2015).

6.4. Berkson Measurement Error

Berkson (1950) introduces a type of measurement error that is complementary to the usual classical assumptions, where the relationship between the correctly measured unobserved variable X^* and the mismeasured observed variable X is a reversed version of the classical errors:

$$X^* = X + \Delta X^*, \tag{90}$$

where ΔX^* is independent of X. This setup has a long history in statistics and is highly plausible in a number of settings. One may attempt to impose a certain value of the control variable (e.g., medication dosage or oven temperature), but the actual value of that variable may deviate randomly from this value. In economics, it is being increasingly recognized that Berkson-type errors may occur, for instance, when the agents reporting the data attempt to form the best possible predictor given their available information (Hyslop & Imbens 2001, Hoderlein & Winter 2010). Another

plausible occurrence is when, for example, only a regional average is observed for one of the regressors, although the researcher is interested in a model at the individual level (Schennach 2013b). It has also been suggested that some measurement errors in the regressors can consist of a combination of a classical and a Berkson error (see, e.g., Hyslop & Imbens 2001).

For reference, a small-variance approximation similar to Chesher (1991), adapted to the Berkson case, is provided here:

$$E\left[Y\,|X\,=x\right]=g\left(x\right)-\frac{\sigma^{2}}{2}g''\left(x\right)+o\left(\sigma^{2}\right), \tag{91}$$

$$f_{X}\left(x\right)=f_{X^{*}}\left(x\right)-\frac{\sigma^{2}}{2}f''_{X^{*}}\left(x\right)+o\left(\sigma^{2}\right). \tag{92}$$

The bias contains terms similar to Equations 7 and 8, with the expected opposite sign. But perhaps more surprisingly, Equation 91 is free of terms that involve first derivatives, which reflects the known fact that a conventional least-squares regression of Y on a Berkson-type error–contaminated regressor X is only consistent for linear or quadratic specifications, with a biased intercept coefficient in the quadratic case.

A popular approach to identification and estimation is to impose parametric restrictions on the distributions of the variables and solve for the parameter values that reproduce various conditional moments of the observed variables. This approach has been used, in particular, for polynomial specifications (Huwang & Huang 2000) and more recently for a very wide range of parametric models (Wang 2004, 2007). Nonparametric estimation, but under the assumption that the distribution of the Berkson error ΔX^{*} is fully known, has also been demonstrated (Delaigle et al. 2006, Carroll et al. 2007).

Schennach (2013b) addresses the question of identification and estimation of a fully nonparametric regression with Berkson measurement error in the regressors of an unknown distribution using an instrument Z related to the true regressor X^{*} via

$$Z=h\left(X^{*}\right)+\Delta Z, \tag{93}$$

where the function $h(\cdot)$ is also unknown. The following assumptions are made: (*a*) X, ΔX^{*}, ΔY, and ΔZ are mutually independent (this can be relaxed, as explained in Schennach 2013b), (*b*) ΔX^{*}, ΔY, and ΔZ are centered (i.e., have zero mean, mode, median, etc.), (*c*) the marginal CFs of ΔX^{*} and ΔZ are nonvanishing, (*d*) the functions $g(x^{*})$ and $h(x^{*})$ are one-to-one, and (*e*) h and its inverse are differentiable. Identification can then be stated as follows: Given the true observed conditional density $f_{Y,Z|X}$, the functional equation

$$f_{Y,Z|X}\left(y,z|x\right)=\int f_{\Delta Z}\left(z-h\left(x^{*}\right)\right)f_{\Delta Y}\left(y-g\left(x^{*}\right)\right)f_{\Delta X^{*}}\left(x^{*}-x\right)\mathrm{d}x^{*} \tag{94}$$

admits a unique solution $(g, h, f_{\Delta Z}, f_{\Delta Y}, f_{\Delta X^{*}})$.

This result is proven using operator techniques analogous to those of Hu & Schennach (2008), but with some crucial differences. First, following steps similar to those in Section 6.2, Equation 94 implies the operator equivalence

$$L_{y;Z|X}L_{Z|X}^{-1}=L_{Z|X^{*}}D_{y;X^{*}}L_{Z|X^{*}}^{-1}. \tag{95}$$

However, one cannot proceed in the same way to ensure that this decomposition is unique because it does not make sense to assume that the distribution of Z given X^{*} is centered in any way, as Z and X^{*} do not even have the same units. Instead, the proof proceeds by showing that a given operator $L_{Z|X}$ corresponds to a unique operator $L_{X^{*}|X}$, via

$$L_{X^{*}|X}=L_{Z|X^{*}}^{-1}L_{Z|X}, \tag{96}$$

and the conditional density $f_{X^*|X}$ (associated with $L_{X^*|X}$) can plausibly be assumed to be centered at $X^* = x$, given $X = x$.

Data subject to rounding errors also follow the structure of Equation 90, although the error ΔX^* is generally not independent from X in that case. Rounding with a systematic rule is best handled with the bounding techniques discussed in the next section. However, the rounding problem acquires a complex nature if rounding is performed by the reporting agent as a reflection of his or her level of confidence in the accuracy of his or her report. In this case, more observations tend to bunch at coarse fractions (such as multiples of 0.5) than at finer ones (such as multiples of 0.1). Methods to handle this require a suitable behavioral model of the respondents (Hoderlein et al. 2015).

6.5. Bounding

If the conditions needed for point identification are too stringent, another way to approach nonclassical measurement error is, not surprisingly, to derive bounds on the parameters of interest under weaker conditions, in the spirit of the literature on set identification (e.g., Manski 1990, 2003; Imbens & Manski 2004; Chernozhukov et al. 2007a). For instance, respondents in a survey may report ranges instead of specific values (i.e., interval-valued data or rounding errors), or some data could be missing or erroneous with some probability (i.e., contaminated or corrupted data).

Examples of this line of work include Manski & Tamer (2002), who derive bounds on the regression coefficients when both the regressors and dependent variable could be interval valued. Additionally, Horowitz & Manski (1995) consider the problem of bounding the true distribution of a variable with a known support, based on contaminated data. Molinari (2008) tackles the general misclassification problem without any auxiliary variables by using a bounding approach and characterizes the identified set under a wide range of plausible assumptions regarding the measurement error (e.g., bounds on the misclassification probabilities, symmetry, and monotonicity constraints).

The ELVIS method (Schennach 2014), introduced in Section 5.2, offers a general and automatic way to handle these general nonclassical data problems, with the additional benefit that it considerably facilitates the inclusion of additional plausible constraints that may help narrow down the identified set, without having to revisit the entire analysis of identification. Schennach (2014, figures 2 and 3) illustrates this idea for interval-valued data and censored regressions, showing that higher-order moment constraints (which impose plausible constraints on the form of heteroskedasticity) help reduce the range of possible parameter values. These higher-order moment constraints involve nonmonotone functions, so an analytic derivation of the bounds would have been daunting. ELVIS is not limited to higher-order moments—it can just as easily handle the median restrictions on the measurement error suggested by Hu & Schennach (2008), for instance, or make use of instruments and repeated measurements, which could also be interval valued or mismeasured, etc.

7. APPLICATIONS TO OTHER FIELDS

Techniques developed for measurement error problems have already found application in various fields. In the identification of auction models, for instance, there is a natural analogy between the measurement error structure and the decomposition of the value of an auctioned item into a sum of a common and private component (Li et al. 2000; Athey & Haile 2002, 2007; An et al. 2010; Krasnokutskaya 2011; Hu et al. 2013). In panel data models with fixed effects, Evdokimov (2009) shows that, upon suitable conditioning, observations in two time periods are analogous to

repeated measurements that can disentangle the distribution of the fixed effect from those of the transitory errors. In random-coefficients panel data models, generalizations of Kotlarski's lemma lead to identification of the distribution of the random coefficients and time-varying error distributions (Arellano & Bonhomme 2011). Treatment effect models in which the control variables only enable imperfect matching can also benefit from measurement error techniques (Heckman et al. 2010a). Measurement error and latent variables techniques have also found applications in game-theoretic settings when a player's actions depend on unobserved variables (Ciliberto & Tamer 2009; Beresteanu et al. 2011; Hu & Shum 2013). Hu (2015) surveys applications of measurement error models in industrial organization and labor economics. General methods that can be used for bounding or point identification in nonlinear measurement error models also have clear applications in general latent variable models (Schennach 2014), such as interval-valued data, limited dependent variables, panel data, and some game-theoretic models.

8. CONCLUSION

This article provides an overview of the field of measurement error in nonlinear models, from simple methods that assume distributional knowledge regarding the measurement error to advanced methods that eliminate measurement error biases via readily available information, even in nonclassical settings.

Where does the future of the field of measurement error modeling lie? Certainly, the field of nonclassical measurement error is still rapidly growing. More closed-form identification results enabling natural plug-in estimators that avoid the need for high-dimensional nonparametric series or sieve estimators would be welcome. Handling the combination of measurement error with other data problems via generic latent variables techniques (e.g., Schennach 2014) in point- or set-identified settings is also a promising area. Finally, existing techniques that properly account for the presence of measurement error should be more broadly used in applications, as rich data sets (potentially with repeated measurements or instruments) are becoming more common. This transition would be facilitated by the creation of easy-to-use, widely available, and robust estimation routines. Ultimately, it would be fruitful if some of the identification results were used as guidance to design collection methods that move away from the elusive goal of capturing exact data and instead focus on the feasible task of obtaining mismeasured data with properties compatible with the most powerful identification results.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGEMENTS

The author acknowledges useful comments by Florian Gunsilius and financial support from NSF grant SES-1357401.

LITERATURE CITED

Abadie A, Angrist JD, Imbens GW. 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70:91–117

Ai C, Chen X. 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71:1795–843

Allman ES, Matias C, Rhodes JA. 2009. Identifiability of parameters in latent structure models with many observed variables. *Ann. Stat.* 37:3099–132

Amemiya Y. 1985. Instrumental variable estimator for the nonlinear errors-in-variables model. *J. Econom.* 28:273–89

An Y, Hu Y. 2012. Well-posedness of measurement error models for self-reported data. *J. Econom.* 168:259–69

An Y, Hu Y, Shum M. 2010. Estimating first-price auctions with an unknown number of bidders: a misclassification approach. *J. Econom.* 157:328–41

Anderson TW, Rubin H. 1956. Statistical inference in factor analysis. *Proc. 3rd Berkeley Symp. Math. Stat. Probab.*, Vol. 5, pp. 111–50. Berkeley: Univ. Calif. Press

Andrews DWK. 2011. *Examples of L^2-complete and boundedly-complete distributions*. Work. Pap. 1801, Cowles Found., Yale Univ., New Haven, CT

Arellano M, Bonhomme S. 2011. Identifying distributional characteristics in random coefficients panel data models. *Rev. Econ. Stud.* 79:987–1020

Arellano M, Bonhomme S. 2015. *Nonlinear panel data estimation via quantile regressions*. Cemmap Work. Pap. CWP40/15, Inst. Fisc. Stud., London

Athey S, Haile PA. 2002. Identification of standard auction models. *Econometrica* 70:2107–40

Athey S, Haile PA. 2007. Nonparametric approaches to auctions. See Heckman & Leamer 2007, pp. 3847–965

Battistin E, Chesher A. 2014. Treatment effect estimation with covariate measurement error. *J. Econom.* 178:707–15

Ben-Moshe D. 2013. *Identification and estimation of factor models using the log characteristic function*. Work. Pap., Hebrew Univ., Jerusalem

Ben-Moshe D. 2014. *Identification of dependent nonparametric distributions in a system of linear equations*. Work. Pap., Hebrew Univ., Jerusalem

Beresteanu A, Molchanov I, Molinari F. 2011. Sharp identification regions in models with convex moment predictions. *Econometrica* 79:1785–821

Berkson J. 1950. Are there two regressions? *J. Am. Stat. Assoc.* 45:164–80

Bollinger CR. 1998. Measurement error in the current population survey: a nonparametric look. *J. Labor Econ.* 16:576–94

Bonhomme S, Jochmans K, Robin JM. 2016. Estimating multivariate latent-structure models. *Ann. Stat.* 44:540–63

Bonhomme S, Robin JM. 2009. Consistent noisy independent component analysis. *J. Econom.* 149:12–25

Bonhomme S, Robin JM. 2010. Generalized non-parametric deconvolution with an application to earnings dynamics. *Rev. Econ. Stud.* 77:491–533

Bound J, Brown C, Mathiowetz N. 2001. Measurement error in survey data. In *Handbook of Econometrics*, Vol. 5, ed. JJ Heckman, E Leamer, pp. 3705–843. Amsterdam: North-Holland

Bound J, Krueger AB. 1991. The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *J. Labor Econ.* 9:1–24

Browning M, Crossley TF. 2009. Are two cheap, noisy measures better than one expensive, accurate one? *Am. Econ. Rev.* 99:99–103

Cameron A, Li T, Trivedi P, Zimmer D. 2004. Modeling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *Econom. J.* 7:566–84

Carrasco M, Florens JP, Renault E. 2007. Linear inverse problems and structural econometrics: estimation based on spectral decomposition and regularization. See Heckman & Leamer 2007, pp. 5633–751

Carroll RJ, Chen X, Hu Y. 2010. Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. *J. Nonparametr. Stat.* 22:379–99

Carroll RJ, Delaigle A, Hall P. 2007. Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *J. R. Stat. Soc. B* 69:859–78

Carroll RJ, Hall P. 1988. Optimal rates of convergence for deconvolving a density. *J. Am. Stat. Assoc.* 83:1184–86

Carroll RJ, Ruppert D, Crainiceanu CM, Tosteson TD, Karagas MR. 2004. Nonlinear and nonparametric regression and instrumental variables. *J. Am. Stat. Assoc.* 99:736–50

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement Error in Nonlinear Models*. New York: Chapman & Hall
Chen X. 2007. Large sample sieve estimation of semi-nonparametric models. See Heckman & Leamer 2007, pp. 5549–632
Chen X, Hong H, Nekipelov D. 2011. Nonlinear models of measurement errors. *J. Econ. Lit.* 49:901–37
Chen X, Hong H, Tamer E. 2005. Measurement error models with auxiliary data. *Rev. Econ. Stud.* 72:343–66
Chen X, Hu Y, Lewbel A. 2009. Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information. *Stat. Sin.* 19:949–68
Cheng CL, Ness JWV. 1999. *Statistical Regression with Measurement Error*. London: Arnold
Chernozhukov V, Hansen C. 2005. An IV model of quantile treatment effects. *Econometrica* 73:245–61
Chernozhukov V, Hong H, Tamer E. 2007a. Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75:1243–84
Chernozhukov V, Imbens GW, Newey WK. 2007b. Instrumental variable identification and estimation of nonseparable models via quantile conditions. *J. Econom.* 139:4–14
Chesher A. 1991. The effect of measurement error. *Biometrika* 78:451–62
Chesher A. 2001. *Parameter approximations for quantile regressions with measurement error*. Work. Pap. CWP02/01, Dep. Econ., Univ. Coll. London
Chesher A. 2003. Identification in nonseparable models. *Econometrica* 71:1405–41
Ciliberto F, Tamer E. 2009. Market structure and multiple equilibria in airline markets. *Econometrica* 77:1791–828
Cragg JC. 1997. Using higher moments to estimate the simple errors-in-variables model. *Rand J. Econ.* 28:S71–91
Cunha F, Heckman J, Schennach SM. 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78:883–931
Dagenais MG, Dagenais DL. 1997. Higher moment estimators for linear regression models with errors in variables. *J. Econom.* 76:193–221
Darolles S, Florens JP, Renault E. 2011. Nonparametric instrumental regression. *Econometrica* 79:1541–65
Delaigle A, Hall P, Meister A. 2008. On deconvolution with repeated measurements. *Ann. Stat.* 36:665–85
Delaigle A, Hall P, Qiu P. 2006. Nonparametric methods for solving the Berkson errors-in-variables problem. *J. R. Stat. Soc. B* 68:201–20
d'Haultfoeuille X. 2011. On the completeness condition in nonparametric instrumental problems. *Econom. Theory* 27:460–71
Dunford N, Schwartz JT. 1971. *Linear Operators*. New York: Wiley
Ebrahimi N, Hamedani G, Soofi ES, Volkmer H. 2010. A class of models for uncorrelated random variables. *J. Multivar. Anal.* 101:1859–71
Ekeland I, Galichon A, Henry M. 2010. Optimal transportation and the falsifiability of incompletely specified economic models. *Econ. Theory* 42:355–74
Erickson T. 1993. Restricting regression slopes in the errors-in-variables model by bounding the error correlation. *Econometrica* 61:959–69
Erickson T, Whited TM. 2000. Measurement error and the relationship between investment and *q*. *J. Polit. Econ.* 108:1027–57
Erickson T, Whited TM. 2002. Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econom. Theory* 18:776–99
Evdokimov K. 2009. *Identification and estimation of a nonparametric panel data model with unobserved heterogeneity*. Work. Pap., Yale Univ., New Haven, CT
Evdokimov K, White H. 2012. Some extensions of a lemma of Kotlarski. *Econom. Theory* 28:925–32
Fan J. 1991a. Asymptotic normality for deconvolution kernel density estimators. *Sankhyā A* 53:97–110
Fan J. 1991b. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Stat.* 19:1257–72
Fan J, Truong YK. 1993. Nonparametric regression with errors in variables. *Ann. Stat.* 21:1900–25
Freyberger J. 2012. *Nonparametric panel data models with interactive fixed effects*. Work. Pap., Northwestern Univ., Evanston, IL

Frisch R. 1934. *Statistical Confluence Analysis by Means of Complete Regression Systems*. Oslo: Univ. Inst. Econ.
Fuller WA. 1987. *Measurement Error Models*. New York: Wiley
Galichon A, Henry M. 2013. Dilation bootstrap: a methodology for constructing confidence regions with partially identified models. *J. Econom.* 177:109–15
Gallant AR, Nychka DW. 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica* 55:363–90
Geary RC. 1942. Inherent relations between random variables. *Proc. R. Irish Acad.* 47A:63–76
Gel'fand IM, Shilov GE. 1964. *Generalized Functions*. New York: Academic
Griliches Z, Hausman JA. 1986. Errors in variables in panel data. *J. Econom.* 31:93–118
Hall P, Horowitz JL. 2005. Nonparametric methods for inference in the presence of instrumental variables. *Ann. Stat.* 33:2904–29
Hamedani GG, Volkmer HW. 2009. Comment on "conditional moments and independence" by A. De Paula. *Am. Stat.* 63:295–95
Hansen LP. 1982. Large sample properties of generalized method of moment estimators. *Econometrica* 50:1029–54
Härdle W, Linton O. 1994. Applied nonparametric methods. In *Handbook of Econometrics*, Vol. 4, ed. R Engle, D McFadden, pp. 2295–339. Amsterdam: North-Holland
Hausman J. 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *J. Econ. Perspect.* 15:57–67
Hausman J, Abrevaya J, Scott-Morton FM. 1998. Misclassification of the dependent variable in a discrete-response setting. *J. Econom.* 87:237–69
Hausman J, Newey W, Ichimura H, Powell J. 1991. Measurement errors in polynomial regression models. *J. Econom.* 50:273–95
Hausman J, Newey W, Powell J. 1995. Nonlinear errors in variables. Estimation of some Engel curves. *J. Econom.* 65:205–33
Heckman JJ, Leamer EE, eds. 2007. *Handbook of Econometrics*, Vol. 6. Amsterdam: North-Holland
Heckman JJ, Schennach SM, Williams BD. 2010a. *Matching on proxy variables*. Work. Pap., Univ. Chicago
Heckman JJ, Schennach SM, Williams BD. 2010b. *Nonparametric factor score regression with an application to the technology of skill formation*. Work. Pap., Univ. Chicago
Heckman JJ, Vytlacil E. 2005. Structural equations, treatment effects and econometric policy evaluation. *Econometrica* 73:669–738
Hoderlein S, Mammen E. 2007. Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75:1513–18
Hoderlein S, Siflinger B, Winter J. 2015. *Identification of structural models in the presence of measurement error due to rounding in survey responses*. Work. Pap., Boston Coll.
Hoderlein S, Winter J. 2010. Structural measurement errors in nonseparable models. *J. Econom.* 157:432–40
Horowitz J, Manski C. 1995. Identification and robustness with contaminated and corrupted data. *Econometrica* 63:281–302
Horowitz JL, Markatou M. 1996. Semiparametric estimation of regression models for panel data. *Rev. Econ. Stud.* 63:145–68
Hsiao C. 1989. Consistent estimation for some nonlinear errors-in-variables models. *J. Econom.* 41:159–85
Hu Y. 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution. *J. Econom.* 144:27–61
Hu Y. 2015. *Microeconomic models with latent variables: applications of measurement error models in empirical industrial organization and labor economics*. Cemmap Work. Pap. CWP03/15, Inst. Fisc. Stud., London
Hu Y, McAdams D, Shum M. 2013. Nonparametric identification of first-price auctions with non-separable unobserved heterogeneity. *J. Econom.* 174:186–93
Hu Y, Ridder G. 2010. On deconvolution as a first stage nonparametric estimator. *Econom. Rev.* 29:1–32
Hu Y, Ridder G. 2012. Estimation of nonlinear models with measurement error using marginal information. *J. Appl. Econom.* 27:347–85
Hu Y, Sasaki Y. 2015. Closed-form estimation of nonparametric models with non-classical measurement errors. *J. Econom.* 185:392–408
Hu Y, Schennach SM. 2008. Instrumental variable treatment of nonclassical measurement error models. *Econometrica* 76:195–216

Hu Y, Schennach SM, Shiu JL. 2015. *Injectivity of a class of integral operators with compactly supported kernels*. Work. Pap., Brown Univ., Providence, RI
Hu Y, Shiu JL. 2013. Identification and estimation of nonlinear dynamic panel data models with unobserved covariates. *J. Econom.* 175:116–31
Hu Y, Shum M. 2012. Nonparametric identification of dynamic models with unobserved state variables. *J. Econom.* 171:32–44
Hu Y, Shum M. 2013. Identifying dynamic games with serially-correlated unobservables. *Adv. Econom.* 31:97–113
Huwang L, Huang YHS. 2000. On errors-in-variables in polynomial regressions—Berkson case. *Stat. Sin.* 10:923–36
Hyslop DR, Imbens GW. 2001. Bias from classical and other forms of measurement error. *J. Bus. Econ. Stat.* 19:475–81
Imbens G, Manski C. 2004. Confidence intervals for partially identified parameters. *Econometrica* 72:1845–57
Imbens GW, Spady RH, Johnson P. 1998. Information theoretic approaches to inference in moment condition models. *Econometrica* 66:333–57
Kapteyn A, Wansbeek T. 1983. Identification in the linear errors in variables model. *Econometrica* 51:1847–49
Kendall MG, Stuart A. 1979. *The Advanced Theory of Statistics*. New York: Macmillan. 4th ed.
Kitamura Y, Stutzer M. 1997. An information-theoretic alternative to generalized method of moment estimation. *Econometrica* 65:861–74
Klepper S, Leamer EE. 1984. Consistent sets of estimates for regressions with errors in all variables. *Econometrica* 52:163–83
Kolda TG, Bader BW. 2009. Tensor decompositions and applications. *SIAM Rev.* 51:455–500
Kotlarski I. 1967. On characterizing the gamma and the normal distribution. *Pac. J. Math.* 20:69–76
Krasnokutskaya E. 2011. Identification and estimation in procurement auctions under unobserved auction heterogeneity. *Rev. Econ. Stud.* 78:293–327
Kruskal JB. 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra Appl.* 18:95–138
Kruskal JB. 1989. Rank, decomposition, and uniqueness for 3-way and *n*-way arrays. In *Multiway Data Analysis*, ed. R Coppi, S Bolasco, pp. 7–18. Amsterdam: North-Holland
Laczkovich M. 1984. Differentiable restrictions of continuous functions. *Acta Math. Hung.* 44:355–60
Lewbel A. 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65:1201–13
Lewbel A. 2007. Estimation of average treatment effects with misclassification. *Econometrica* 75:537–51
Lewbel A. 2012. Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. *J. Bus. Econ. Stat.* 30:67–80
Li T. 2002. Robust and consistent estimation of nonlinear errors-in-variables models. *J. Econom.* 110:1–26
Li T, Perrigne I, Vuong Q. 2000. Conditionally independent private information in OCS wildcat auctions. *J. Econom.* 98:129–61
Li T, Trivedi P, Guo J. 2003. Modeling response bias in count: a structural approach with an application to the national crime victimization survey data. *Sociol. Methods Res.* 31:514–44
Li T, Vuong Q. 1998. Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivar. Anal.* 65:139–65
Lighthill MJ. 1962. *Introduction to Fourier Analysis and Generalized Function*. Cambridge, UK: Cambridge Univ. Press
Liu M, Taylor R. 1989. A consistent nonparametric density estimator for the deconvolution problem. *Can. J. Stat.* 17:427–38
Loève M. 1977. *Probability Theory I*. New York: Springer
Lukacs E. 1970. *Characteristic Functions*. London: Griffin. 2nd ed.
Mahajan A. 2006. Identification and estimation of single index models with misclassified regressor. *Econometrica* 74:631–65
Manski C. 1990. Nonparametric bounds on treatment effects. *Am. Econ. Rev.* 80:319–23
Manski C. 2003. *Partial Identification of Probability Distributions*. New York: Springer-Verlag

Manski CF, Tamer E. 2002. Inference on regressions with interval data on a regressor or outcome. *Econometrica* 70:519–46

Mattner L. 1993. Some incomplete but boundedly complete location families. *Ann. Stat.* 21:2158–62

Matzkin RL. 2003. Nonparametric estimation of nonparametric nonadditive random functions. *Econometrica* 71:1339–75

Matzkin RL. 2008. Identification in nonparametric simultaneous equations. *Econometrica* 76:945–78

McFadden D. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57:995–1026

McIntyre J, Stefanski LA. 2011. Heteroscedastic measurement deconvolution. *Ann. Inst. Stat. Math.* 63:81–99

Molinari F. 2008. Partial identification of probability distributions with misclassified data. *J. Econom.* 144:81–117

Nadai MD, Lewbel A. 2016. Nonparametric errors in variables models with measurement errors on both sides of the equation. *J. Econom.* 191:19–32

Newey WK. 2001. Flexible simulated moment estimation of nonlinear errors-in-variables models. *Rev. Econ. Stat.* 83:616–27

Newey WK, Powell JL. 2003. Instrumental variable estimation of nonparametric models. *Econometrica* 71:1565–78

Newey WK, Smith RJ. 2004. Higher-order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72:219–55

Owen AB. 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75:237–49

Owen AB. 1990. Empirical likelihood ratio confidence regions. *Ann. Stat.* 18:90–120

Pakes A, Pollard D. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57:1027–57

Pal M. 1980. Consistent moment estimators of regression coefficients in the presence of errors in variables. *J. Econom.* 14:349–64

Qin J, Lawless J. 1994. Empirical likelihood and general estimating equations. *Ann. Stat.* 22:300–25

Rao P. 1992. *Identifiability in Stochastic Models*. New York: Academic

Reiersol O. 1950. Identifiability of a linear relation between variables which are subject to error. *Econometrica* 18:375–89

Ridder G, Moffitt R. 2007. The econometrics of data combination. See Heckman & Leamer 2007, pp. 5469–547

Sasaki Y. 2015. Heterogeneity and selection in dynamic panel data. *J. Econom.* 188:236–49

Schennach SM. 2000. *Estimation of nonlinear models with measurement error*. Work. Pap., Univ. Chicago

Schennach SM. 2004a. Estimation of nonlinear models with measurement error. *Econometrica* 72:33–75

Schennach SM. 2004b. Exponential specifications and measurement error. *Econ. Lett.* 85:85–91

Schennach SM. 2004c. Nonparametric estimation in the presence of measurement error. *Econ. Theory* 20:1046–93

Schennach SM. 2007a. Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica* 75:201–39

Schennach SM. 2007b. Point estimation with exponentially tilted empirical likelihood. *Ann. Stat.* 35:634–72

Schennach SM. 2008. Quantile regression with mismeasured covariates. *Econom. Theory* 24:1010–43

Schennach SM. 2013a. *Convolution without independence*. Cemmap Work. Pap. WP46/13, Inst. Fisc. Stud., London

Schennach SM. 2013b. Instrumental variable treatment of the nonparametric Berkson measurement error model. *Ann. Stat.* 41:1642–68

Schennach SM. 2013c. Measurement error in nonlinear models—a review. In *Advances in Economics and Econometrics*, Vol. 3, ed. D Acemoglu, M Arellano, E Dekel, pp. 296–337. Cambridge, UK: Cambridge Univ. Press

Schennach SM. 2014. Entropic latent variable integration via simulation. *Econometrica* 82:345–86

Schennach SM, Hu Y. 2013. Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *J. Am. Stat. Assoc.* 108:177–86

Schennach SM, White H, Chalak K. 2012. Local indirect least squares and average marginal effects in nonseparable structural systems. *J. Econom.* 166:282–302

Schwartz L. 1966. *Théorie des distributions*. Paris: Hermann
Sepanski JH, Carroll RJ. 1993. Semiparametric quasi-likelihood and variance function estimation in measurement error models. *J. Econom.* 58:223–56
Shen X. 1997. On methods of sieves and penalization. *Ann. Stat.* 25:2555–91
Song S, Schennach S, White H. 2015. Estimating nonseparable models with mismeasured endogenous variables. *Quant. Econ.* 6:749–94
Stegeman A. 2009. On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode. *Linear Algebra Appl.* 431:211–27
Taupin ML. 2001. Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Stat.* 29:66–93
Temple G. 1963. The theory of weak functions. I. *Proc. R. Soc. Lond. A* 276:149–67
Wang L. 2004. Estimation of nonlinear models with Berkson measurement errors. *Ann. Stat.* 32:2559–79
Wang L. 2007. A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models. *Can. J. Stat.* 35:233–48
Wang L, Hsiao C. 2011. Method of moments estimation and identifiability of nonlinear semiparametric errors-in-variables models. *J. Econom.* 165:30–44
Wansbeek TJ, Meijer E. 2000. *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier
Wilhelm D. 2010. *Identification and estimation of nonparametric panel data regressions with measurement error*. Work. Pap., Univ. Chicago