

Measurement error in nonlinear models- a review

S. M. Schennach

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP41/12

Measurement Error in Nonlinear Models — A Review

Susanne M. Schennach

November 15, 2012

Abstract

This overview of the recent econometrics literature on measurement error in nonlinear models centers on the question of the identification and estimation of general nonlinear models with measurement error. Simple approaches that rely on distributional knowledge regarding the measurement error (such as deconvolution or validation data techniques) are briefly presented. Then follows a description of methods that secure identification via more readily available auxiliary variables (such as repeated measurements, measurement systems with a “factor model” structure, instrumental variables and panel data). Methods exploiting higher-order moments or bounding techniques to avoid the need for auxiliary information are 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 active topic of nonclassical measurement error is covered and applications of measurement error techniques to other fields are outlined.

Forthcoming in:

Advances in Economics and Econometrics: Theory and Applications,

Tenth World Congress, Edited by D. Acemoglu, M. Arellano and E. Dekel.

1 Introduction

Measurement error is widespread in statistical and/or economic data and can have substantial impact on point estimates and statistical inference in general. Accordingly there exists a vast literature focused on addressing this problem. The present overview emphasizes the recent econometric literature on the topic and mostly centers on the author’s interest in the question of identification (and consistent estimation) of general nonlinear models with measurement error *without* simply assuming that the distribution of the measurement error is known.

This paper is organized as follows. First, we explain the origins of measurement error bias, before describing simple approaches that rely on distributional knowledge regarding the measurement error (such as deconvolution or validation data techniques). We then describe methods that secure identification via more readily available auxiliary variables (such as repeated measurements, 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 Models and conventions

Let us first summarize the models and conventions used throughout. The unobserved but true value of the variable of interest X^* and its observed but mismeasured counterpart X are related through

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

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 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 (such as instruments) that will be denoted by Z . When the model includes disturbances in the process generating Y or Z , these will be denoted by ΔY and ΔZ , respectively.

A lower case letter denotes a specific value of the random variable (or random vector) denoted by the corresponding upper-case letter. Let $f_V(v)$ and $F_V(v)$ denote, respectively, the density¹ and the distribution of the random variable V and similarly for random vectors, whose dimension is denoted d_V . For the most part, we shall omit regularity conditions and focus on the substantive assumptions needed for identification. On occasion, some models will be a simplified version of the original works to streamline the presentation. All 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 measurement error literature has considered the so-called *classical measurement error* assumptions, which may have two alternative meanings depending on the context:

1. ΔX independent from X^* and $E[\Delta X] = 0$ (which could be called “strongly classical”) or
2. $E[\Delta X|X^* = x^*] = 0$ (which could be called “weakly classical”, as it allows for heteroskedasticity in the measurement error).

Depending on the specific model, these assumptions may include independence from or conditioning on some of the other variables of the model ($Y, \Delta Y, Z, \Delta Z$). Deviations from these assumptions would be labelled *nonclassical measurement error*, which is currently the focus of intense and increasing attention.

This review focuses on nonlinear models (as a linear model with classical measurement errors can be straightforwardly handled via standard instrumental variable techniques). Examples include:

1. Nonlinear regression models of the form

$$Y = g(X^*, \theta) + \Delta Y \tag{2}$$

with $E[\Delta Y|X^* = x^*] = 0$ and where the nonlinearity here refers to nonlinearity in X^* , which is the real source of the identification challenge — nonlinearity in the parameter vector θ can be handled through standard asymptotic techniques. The regression function $g(X^*, \theta)$ can be either parametric or nonparametric.

¹We work mostly with densities (with respect to the Lebesgue measure) throughout, but most identification results hold for general probability measures, at the expense of notational complications.

2. Maximum likelihood models

$$\theta = \arg \max_{\theta} E [\ln f_{Y, X^* | \theta} (Y, X^* | \theta)] \quad (3)$$

where the likelihood function $f_{Y, X^* | \theta} (y, x^* | \theta)$ is parametrically specified via a parameter vector θ .

3. Generalized method of moment (GMM) models (Hansen (1982))

$$E [m (Y, X^*, \theta)] = 0 \quad (4)$$

where the moment function $m (y, x^*, \theta)$ is parametrically specified via a parameter vector θ .

4. More broadly, we consider the problem of recovering the unobserved density $f_{X^*} (x^*)$ or, more generally, the unobserved joint density $f_{Y, X^*} (y, x^*)$ from densities of observable variables. From the latter, one can, in principle, identify any quantity of interest, including the conditional expectation $g (x^*) \equiv E [Y | X^* = x^*]$.

2 The effect of measurement error

Most econometrics textbooks emphasize the fact that, in a simple linear bivariate regression, the presence of measurement error “attenuates” the relationship between the dependent variable and the mismeasured regressors. This result 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 (2001), Hausman, Newey, and Powell (1995), Griliches and Ringstad (1970)). 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 the 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, $E [a (X^* + \Delta X)] \neq E [a (X^*)]$ for a nonlinear function $a (\cdot)$ (which could be part of the model of interest or an estimator of that model, for instance). 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

$$E[Y|X = x] = g(x) + \frac{\sigma^2}{2} (g''(x) + 2g'(x) (\ln f_{X^*}(x))') + o(\sigma^2) \quad (5)$$

$$f_X(x) = f_{X^*}(x) + \frac{\sigma^2}{2} f_{X^*}''(x) + o(\sigma^2) \quad (6)$$

where primes denote derivatives. He also shows that one can use the approximations $g(x) \approx E[Y|X = x]$ and $f_{X^*}(x) \approx f_X(x)$ in (5) and (6) 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 ($(\ln f_{X^*}(x))'$), while 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 and Chesher (2009)).

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

We first introduce 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 (c.f.). The c.f. of a random vector X taking value in \mathbb{R}^{d_X} is defined as

$$\phi_X(\xi) \equiv E \left[e^{\mathbf{i}\xi \cdot X} \right] \quad (7)$$

for all $\xi \in \mathbb{R}^{d_X}$ (where $\mathbf{i} \equiv \sqrt{-1}$) and has some remarkable properties (see Lukacs (1970) and Loève (1977), Sections 13 and 14):

1. There is a one-to-one relationship between $\phi_X(\xi)$ and the probability measure of X ;
2. $\phi_X(\xi)$ always exists, is continuous and bounded (by 1) everywhere;
3. If $X = X^* + \Delta X$ with X^* and ΔX independent, then $\phi_X(\xi) = \phi_{X^*}(\xi) \phi_{\Delta X}(\xi)$.

It is this last property, known as the convolution theorem, which is particularly useful for classical measurement error problems. If the distribution of the measurement error is known, then $\phi_{\Delta X}(\xi)$ is known, and we can express the c.f. of the true unobserved variables in terms of known ($\phi_{\Delta X}(\xi)$) or observed ($\phi_X(\xi)$) quantities:

$$\phi_{X^*}(\xi) = \frac{\phi_X(\xi)}{\phi_{\Delta X}(\xi)}. \quad (8)$$

One can then recover the density of X^* via an inverse Fourier transform:²

$$f_{X^*}(x^*) = (2\pi)^{-d_X} \int \phi_{X^*}(\xi) e^{-\mathbf{i}\xi \cdot x^*} d\xi \quad (9)$$

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

$$\hat{\phi}_X(\xi) = \dot{\phi}_X(\xi) \phi_K(h\xi) \quad (10)$$

where

$$\dot{\phi}_X(\xi) = \frac{1}{n} \sum_{j=1}^n e^{\mathbf{i}\xi \cdot X_j} \quad (11)$$

is the *empirical c.f.* of the sample (X_1, \dots, X_n) and $\phi_K(\xi)$ is the Fourier transform of the kernel and h is the bandwidth. The empirical c.f. is a common and convenient estimation tool for Fourier-

²We state the result for densities for simplicity, a similar result holds for general probability measures, see Loève (1977) Section 14.3.

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 (see Li and Vuong (1998), Schennach (2004a), Hu and Ridder (2010), Bonhomme and Robin (2010)).

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

$$\hat{f}_{X^*,h}(x^*) \equiv (2\pi)^{-d_X} \int \dot{\phi}_X(\xi) \frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} e^{-i\xi \cdot x^*} d\xi \quad (12)$$

which is the focus of an extensive literature (Fan (1991b) Fan (1991a) Carroll and Hall (1988) Liu and Taylor (1989) McIntyre and Stefanski (2011), among others).

To ensure that the integral in (12) exists, the kernel is selected so that its Fourier transform $\phi_K(\cdot)$ is compactly supported, 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 (12) from its true value $\phi_{X^*}(\xi)$ as

$$\dot{\phi}_X(\xi) \frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} - \phi_{X^*}(\xi) = \left(\dot{\phi}_X(\xi) - \phi_X(\xi) \right) \frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} + \phi_{X^*}(\xi) (\phi_K(h\xi) - 1). \quad (13)$$

A small h leads to a small bias (second term; note that $\phi_K(0) = 1$), but to a larger support of $\phi_K(h\xi)$ and a resulting 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 the 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 (13) 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 frequency $|\xi| \rightarrow \infty$, which, in turn, is related to the smoothness of the corresponding original functions. Smoothness of $f_{X^*}(x^*)$ 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 (13).

In the deconvolution literature, smoothness is traditionally characterized by bounds on the rate of decay of the tail of a c.f. There are two broad typical behaviors:

1. 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., Theorem 3 in Schennach (2004a), for an upper bound on the c.f. and Theorem 1 in Hu and Ridder (2010) for a lower bound.³)

2. 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 Lemma S.2 in Schennach (2007a).)

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 (mean square) convergence rate is of the form $n^{-\eta}$ for some $\eta > 0$. Furthermore, when the density of ΔX is supersmooth, while 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). It turns out that 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 Schennach (2004a), Taupin (2001), Schennach, White, and Chalak (2012), Hu and Ridder (2010)), with either ordinary smooth or supersmooth measurement error distributions.

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

$$\hat{g}_h(x^*) = \frac{\int \frac{\dot{\phi}_X^Y(\xi)}{\dot{\phi}_{\Delta X}(\xi)} \phi_K(h\xi) e^{-i\xi \cdot x^*} d\xi}{\int \frac{\dot{\phi}_X(\xi)}{\dot{\phi}_{\Delta X}(\xi)} \phi_K(h\xi) e^{-i\xi \cdot x^*} d\xi} \quad (14)$$

where we have introduced a quantity related to the empirical c.f.

$$\dot{\phi}_X^Y(\xi) \equiv \sum_{j=1}^n Y_j e^{i\xi \cdot X_j} \quad (15)$$

which can be viewed as an estimator of $E[Y e^{i\xi \cdot X}]$ or, equivalently, of $\int E[Y|X=x] e^{i\xi \cdot x} dF(x)$ or $[\partial \phi_{YX}(\zeta, \xi) / \partial \zeta]_{\zeta=0}$.

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

3.2 Validation data

For a main sample containing data on Y and X (where X is mismeasured), validation data (e.g., Sepanski and Carroll (1993)) typically consists 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 and Moffitt (2006).) 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 (2), via the equality

$$E[Y|X] = \int E[Y|X = x, X^* = x^*] f_{X^*|X}(x^*|x) dx^* = \int E[Y|X^* = x^*] f_{X^*|X}(x^*|x) dx^* \quad (16)$$

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

$$\tilde{g}(x, \theta) \equiv \int g(x^*, \theta) f_{X^*|X}(x^*|x) dx^* \quad (17)$$

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

$$Y = \tilde{g}(X, \theta) + \Delta Y \quad (18)$$

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

The main advantage of validation data is that it offers 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 and Krueger (1991) is 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 below).

Hu and Ridder (2012) suggest an interesting setup in the context of classical measurement error, where the main sample contains Y and X , while the validation sample only contains 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 be summarized as follows: The

measurement error distribution is obtained by deconvoluting 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 (14)).

4 Methods based on auxiliary variables

4.1 Repeated Measurements

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

$$X = X^* + \Delta X \tag{19}$$

$$Z = X^* + \Delta Z \tag{20}$$

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

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), or p. 21 in Rao (1992)). This Lemma (proven more generally below) states that if X and Z take value in \mathbb{R} and $X^*, \Delta X, \Delta Z$ are mutually independent with $E[\Delta X] = 0$ then (provided $E[e^{i\zeta Z}] \neq 0$),

$$\phi_{X^*}(\xi) = \exp\left(\int_0^\xi \frac{E[\mathbf{i}X e^{i\zeta Z}]}{E[e^{i\zeta Z}]} d\zeta\right) \tag{21}$$

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

In the very special case where the distributions of ΔX and ΔZ are identical and symmetric about zero, one can obtain the c.f. of X^* in a simpler way (see Horowitz and Markatou (1996), Li and Vuong (1998), Delaigle, Hall, and Meister (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).

Li and Vuong (1998) prove the consistency of a nonparametric density estimator based on Kotlarski's Lemma. This result is extended by Li (2002) to multivariate settings⁴ (assuming mutual

⁴Both Li and Vuong (1998) and Li (2002) require the observed variable Z to have a nonvanishing c.f. 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.

independence of the elements of both ΔX and ΔZ) and used to correct for measurement error in a nonlinear regression model (2). The idea is to first use the following identity:

$$\tilde{g}(x, \theta) \equiv E[Y|X = x] = \int g(x^*, \theta) f_{X^*|X}(x^*|x) dx^* = \int g(x^*, \theta) \frac{f_{\Delta X}(x - x^*) f_{X^*}(x^*)}{f_X(x)} dx^* \quad (22)$$

where the densities $f_{\Delta X}$ and f_{X^*} can be nonparametrically estimated 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 analogue of $E[(Y - \tilde{g}(X, \theta))^2]$.

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

$$Y = \sum_{j=0}^J \theta_j (X^*)^j + \Delta Y \quad (23)$$

where X^* and Y are scalars and where the disturbance satisfy 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, $\theta \equiv (\theta_0, \dots, \theta_J)$ would be identified from

$$\theta = (E[VV'])^{-1} (E[VY]), \text{ where } V = (1, X^*, \dots, (X^*)^J)' \quad (24)$$

where the elements of $E[VV']$ have the general form $\zeta_m = E[(X^*)^m]$, $m = 0, \dots, 2J$ while elements of $E[VY]$ have the general form $\xi_l = E[Y(X^*)^l]$, $l = 0, \dots, J$. 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:

$$\zeta_j = E[XZ^{j-1}] - \sum_{l=0}^{j-2} \binom{j-1}{l} \zeta_{l+1} \nu_{j-l-1} \quad (25)$$

$$\nu_j = E[Z^j] - \sum_{l=1}^j \binom{j}{l} \zeta_l \nu_{j-l} \quad (26)$$

$$\xi_j = E[YZ^j] - \sum_{l=0}^{j-1} \binom{j}{l} \xi_l \nu_{j-l} \quad (27)$$

Starting from $v_0 = 1$ and $\zeta_1 = E[X]$, the relations (25) and (26) can be used to recursively identify $\nu_1, \zeta_2, \nu_2, \zeta_3, \dots, \nu_{2J-1}, \zeta_{2J}$. Then, ξ_0, \dots, ξ_J are obtained from (27).

Hausman, Newey, and Powell (1995) go a step further and observe that, since polynomials form a basis for functions in L_2 , 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

$$\hat{\theta} = \arg \min_{\theta} \int (g(x^*, \theta) - \hat{g}(x^*))^2 w(x^*) dx^* \quad (28)$$

for some given weighting function $w(x^*)$.

In the case of general nonlinear regressions, both Li (2002) and Hausman, Newey, and Powell (1995) only establish consistency (without a rate or a 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, Newey, and Powell (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, Newey, Ichimura, and Powell (1991). Since a large class of models can be identified from the knowledge of a set of moments, she phrases the result 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, but also for likelihood models. As a special case, for nonlinear regression models (2) 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).

We now show how the repeated measurements can be used to identify such moments (in which X^*, X, Y, Z could be random vectors, with $d_{X^*} = d_X = d_Z$). If (i) $E[\Delta X|X^*, \Delta Z] = 0$, (ii) $E[Y|X^*, \Delta Z] = E[Y|X^*]$,⁵ (iii) ΔZ is independent from X^* , then, for any function $u(x^*)$ with Fourier transform $\mu(\xi)$,

$$E[Yu(X^*)] = (2\pi)^{-d_X} \int_{\mathbb{R}^{d_X}} \mu(-\xi) \frac{E[Ye^{i\xi \cdot Z}]}{E[e^{i\xi \cdot Z}]} \phi_{X^*}(\xi) d\xi \quad (29)$$

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

where the integral is over the whole \mathbb{R}^{d_x} space and the c.f. of X^* is given by

$$\phi_{X^*}(\xi) = \exp \left(\int_0^\xi \frac{E[\mathbf{i}X e^{\mathbf{i}\zeta \cdot Z}]}{E[e^{\mathbf{i}\zeta \cdot Z}]} \cdot d\zeta \right) \quad (30)$$

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 (30) can be written as:

$$\begin{aligned} \frac{E[\mathbf{i}X e^{\mathbf{i}\zeta \cdot Z}]}{E[e^{\mathbf{i}\zeta \cdot Z}]} &= \frac{E[\mathbf{i}X^* e^{\mathbf{i}\zeta \cdot (X^* + \Delta Z)}] + E[\mathbf{i}\Delta X e^{\mathbf{i}\zeta \cdot (X^* + \Delta Z)}]}{E[e^{\mathbf{i}\zeta \cdot (X^* + \Delta Z)}]} \\ &= \frac{E[\mathbf{i}X^* e^{\mathbf{i}\zeta \cdot (X^* + \Delta Z)}] + E[\mathbf{i}E[\Delta X | X^*, \Delta Z] e^{\mathbf{i}\zeta \cdot (X^* + \Delta Z)}]}{E[e^{\mathbf{i}\zeta \cdot (X^* + \Delta Z)}]} \\ &= \frac{E[\mathbf{i}X^* e^{\mathbf{i}\zeta \cdot X^*}] E[e^{\mathbf{i}\zeta \cdot \Delta Z}]}{E[e^{\mathbf{i}\zeta \cdot X^*}] E[e^{\mathbf{i}\zeta \cdot \Delta Z}]} = \frac{E[\mathbf{i}X^* e^{\mathbf{i}\zeta \cdot X^*}]}{E[e^{\mathbf{i}\zeta \cdot X^*}]} = \nabla_\zeta \ln E[e^{\mathbf{i}\zeta \cdot X^*}] \end{aligned} \quad (31)$$

where we have used, in turn, the definition of the repeated measurements, iterated expectations, the conditional mean and independence assumptions regarding the errors (ΔX and ΔZ , respectively), various cancellations and the chain rule. Next, (30) can be shown by noting that the path integral of a gradient yields the original function:

$$\exp \left(\int_0^\xi \nabla_\zeta \ln E[e^{\mathbf{i}\zeta \cdot X^*}] \cdot d\zeta \right) = \exp \left(\ln E[e^{\mathbf{i}\xi \cdot X^*}] - \ln E[e^{\mathbf{i}0 \cdot X^*}] \right) = E[e^{\mathbf{i}\xi \cdot X^*}] = \phi_{X^*}(\xi). \quad (32)$$

The integrand in (29) can then be written as:

$$\begin{aligned} \frac{E[Y e^{\mathbf{i}\xi \cdot Z}]}{E[e^{\mathbf{i}\xi \cdot Z}]} E[e^{\mathbf{i}\xi \cdot X^*}] &= \frac{E[E[Y | X^*, \Delta Z] e^{\mathbf{i}\xi \cdot (X^* + \Delta Z)}]}{E[e^{\mathbf{i}\xi \cdot (X^* + \Delta Z)}]} E[e^{\mathbf{i}\xi \cdot X^*}] \\ &= \frac{E[E[Y | X^*] e^{\mathbf{i}\xi \cdot X^*}] E[e^{\mathbf{i}\xi \cdot \Delta Z}]}{E[e^{\mathbf{i}\xi \cdot X^*}] E[e^{\mathbf{i}\xi \cdot \Delta Z}]} E[e^{\mathbf{i}\xi \cdot X^*}] = E[E[Y | X^*] e^{\mathbf{i}\xi \cdot X^*}] \\ &= \int E[Y | X^* = x^*] f_{X^*}(x^*) e^{\mathbf{i}\xi \cdot x^*} dx^* \equiv \alpha(\xi) \end{aligned} \quad (33)$$

where we have used, in turn, iterated expectations, the assumption regarding Y and ΔZ and various cancellations. Equation (29) is then obtained by using Parseval's identity $(2\pi)^{-d_x} \int \mu(-\xi) \alpha(\xi) d\xi = \int u(x^*) a(x^*) dx^*$, where $a(x^*) = E[Y | X^* = x^*] f_{X^*}(x^*)$ with Fourier transform $\alpha(\xi)$.

⁶We do so without dwelling on technical issues such as 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) can be included in Y if they enter the moments linearly. If Y is not needed, it can be simply set to $Y = 1$. 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), Schwartz (1966), Temple (1963), Gel'fand and Shilov (1964)). 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 $(E[Y e^{i\xi \cdot Z}] / E[e^{i\xi \cdot Z}]) \phi_{X^*}(\xi)$ in (29). In this fashion, one can recover the polynomial result of Hausman, Newey, Ichimura, and Powell (1991) under the same conditional mean and independence assumptions.

Equations (29) and (30) also suggest a very natural estimator in which all quantities of the form $E[V e^{i\xi \cdot Z}]$ for $V = 1, Y$ are replaced by sample averages $n^{-1} \sum_{j=1}^n V_j e^{i\xi \cdot Z_j}$. Schennach (2004a) shows that, after a simple automatic bounding device (ensuring, for instance, that an estimated c.f. obtained via (30) is bounded), 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 $|\xi| \rightarrow \infty$ to downweigh the noise in the tail in the estimated c.f., yielding a finite overall noise that decays to zero at the rate $n^{-1/2}$ as $n \rightarrow \infty$. Schennach, White, and Chalak (2012) extends these results to general semiparametric functionals of densities, conditional expectations and derivatives thereof.

One can also recover Kotlarski's identity from (29) and (30) under fewer independence assumptions by setting $Y = 1$ and $\mu(\xi) = e^{i\xi \cdot x_0^*}$ (which corresponds to setting $u(x^*)$ to be a delta function, or a point mass, at x_0^*). In a similar vein, only setting $\mu(\xi) = e^{i\xi \cdot x_0^*}$ (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 $E[V e^{i\xi \cdot Z}]$ by the corresponding sample average in (29) and (30), because this

yields a noise of infinite magnitude, as $\mu(\xi)$ does not decay as $|\xi| \rightarrow \infty$. As shown in Schennach (2004c), this problem can be fixed 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 $\mu(\xi) = \phi_K(h\xi) e^{i\xi \cdot x_0^*}$ where, 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 $E[V e^{i\xi \cdot Z}]$ for $V = 1, Y$ by $n^{-1} \sum_{j=1}^n V_j e^{i\xi \cdot Z_j}$ in:

$$\frac{\int \phi_K(h\xi) e^{-i\xi \cdot x_0^*} \frac{E[Y e^{i\xi \cdot Z}]}{E[e^{i\xi \cdot Z}]} \phi_{X^*}(\xi) d\xi}{\int \phi_K(h\xi) e^{-i\xi \cdot x_0^*} \phi_{X^*}(\xi) d\xi} \quad (34)$$

with $\phi_{X^*}(\xi)$ given by (30), if one lets $h \rightarrow 0$ at a suitable rate as $n \rightarrow \infty$. 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 where the distribution of the measurement error is fully known. These results are generalized in Schennach, White, and Chalak (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, in order to relax some of the nonvanishing c.f. assumptions, Evdokimov (2009) assumes mutual independence of ΔX , ΔZ , 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 c.f. but not X^* . This can be even further relaxed by evaluating an appropriate limit, provided higher order derivatives of the c.f. do not vanish where the c.f. itself does. This idea is exploited in Schennach (2000) and Evdokimov and White (2011).

4.2 Factor models

It is also possible to extend (19)-(20) to a more general "factor model" structure between the observed measurements X and the unobserved *factors* X^* :

$$X = \Lambda X^* + \Delta X \quad (35)$$

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 and 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 (35), 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, Heckman, and Schennach (2010)). Heckman, Schennach, and Williams (2010) make the important observation 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, using higher order moments can reduce that number (Bonhomme and Robin (2009)), 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 (29)-(30) in order to estimate the joint distribution of the factors X^* (Schennach and Williams (2010)). 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 (35). Then, the repeated measurement vectors X and Z in Equations (29)-(30) 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 and Robin (2010) observe that it is possible to recover the distribution of X^* using even fewer measurements than via a Kotlarski-type identity by working with second order derivatives instead of the first order derivatives corresponding to $E[\mathbf{i}X e^{i\zeta Z}]$ in Kotlarski’s identity (21). Intuitively, this is possible because

second derivatives provide “more equations” than first derivatives.

A generalization (under suitable normalizations) to fully nonlinear nonseparable factor models (i.e. $X = a(X^*, \Delta X)$, for some nonlinear function $a(\cdot, \cdot)$) is developed in Cunha, Heckman, and Schennach (2010), using and extending some of the techniques found in Hu and Schennach (2008) (discussed in more detail in Section 6.2 below).

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 and Crossley (2009)).

4.3 Instrumental Variables

Instrumental variables typically consist of a vector Z of random variables

1. that are related in some way to the true unobserved variable X^* , for instance, through some relationship of the form

$$X^* = h(Z) + \Delta Z \tag{36}$$

where $h(Z)$ is some linear or nonlinear function and

2. that satisfy some exclusion restrictions, such as $E[\Delta X|Z] = 0$ and/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, Newey, Ichimura, and Powell (1991) considers a general polynomial model (slightly more general than reported here) with scalar Y and X^* and with an instrument equation of the form (36) 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 . They first note the function $h(\cdot)$ in (36) is identified from $E[X|Z] = E[X^*|Z] = h(Z)$. In fact, for identification purposes we can simply use $h(Z)$ as the instrument and hence we 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:

$$E[Y|Z = z] = \sum_{j=0}^J \gamma_j z^j \quad (37)$$

$$E[XY|Z = z] = \sum_{j=0}^{J+1} \beta_j z^j \quad (38)$$

The identified coefficients can then be used in a recursive relation to identify θ :

1. Start with $v_0 = 1$, $v_1 = 0$, $\theta_J = \gamma_J$, $\theta_{J-1} = \gamma_{J-1}$.
2. For $k > 1$:

$$v_j = \left(\binom{J}{J-j+1} \theta_J \right)^{-1} \left(\beta_{J-j+1} - \gamma_{J-j} - \sum_{l=J-j+1}^{J-1} \binom{l}{J-j+1} \theta_l v_{l-J+j} \right) \quad (39)$$

$$\theta_{J-j} = \gamma_{J-j} - \sum_{l=J-j+1}^J \binom{l}{J-j} \theta_l v_{l-J+j}. \quad (40)$$

These Equations can be used to recursively find $v_2, \theta_{J-2}, v_3, \theta_{J-3}, \dots, v_J, \theta_0$. 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

$$Y = g(X^*, \theta) + \Delta Y \quad (41)$$

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

$$E[Y|Z = z] = \int g(z - u, \theta) dF_U(u) \quad (42)$$

$$E[XY|Z = z] = \int (z - u) g(z - u, \theta) dF_U(u) \quad (43)$$

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 $dF_U(u)$, 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 (42) and (43), 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 are, in fact, not identified from (42) and (43) 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.

Wang and Hsiao (2011) use (42) and (43) to show how, in a specification $g(x^*, \theta)$ with a d_{X^*} -dimensional x^* , 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 $\int |g(x^*, \theta)| dx^* < \infty$, 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 (42) and (43):

$$\varepsilon_Y(\zeta) = \gamma(\zeta) \phi(\zeta) \quad (44)$$

$$\mathbf{i}\varepsilon_{XY}(\zeta) = \dot{\gamma}(\zeta) \phi(\zeta) \quad (45)$$

where overdots denote derivatives and, with $V = Y$ or XY ,

$$\varepsilon_V(\zeta) = \int E[V|Z=z] e^{\mathbf{i}\zeta z} dz \quad (46)$$

$$\gamma(\zeta) = \int g(x^*) e^{\mathbf{i}\zeta x^*} dx^* \quad (47)$$

$$\phi(\zeta) = E[e^{\mathbf{i}\zeta U}] = E[e^{-\mathbf{i}\zeta \Delta Z}]. \quad (48)$$

These identities can then be manipulated to secure nonparametric identification of $g(x^*)$, under the following conditions: (i) $E[|\Delta Z|] < \infty$, while $|g(x^*)|$, $|E[Y|Z=z]|$, and $|E[XY|Z=z]|$ are bounded by polynomials, and (ii) $\phi(\zeta)$ is nonvanishing and $\gamma_o(\zeta)$ is nonvanishing almost everywhere,⁷ where $\gamma_o(\zeta)$ denotes the *ordinary function* component⁸ of $\gamma(\zeta)$; this allows for the fact that

⁷There is a misprint in the paper, where it should be stated that $\gamma_o(\zeta) \neq 0$ instead of $\gamma(\zeta) \neq 0$. The 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 $|\zeta| > \bar{\zeta}$.

⁸For instance, the Fourier transform of the cdf of a standard normal is $\gamma(\zeta) = \pi\delta(\zeta) - \exp(-\zeta^2/2) / (\mathbf{i}\zeta)$, whose ordinary component is $\gamma_o(\zeta) = -\exp(-\zeta^2/2) / (\mathbf{i}\zeta)$ while $\delta(\zeta)$ denotes 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

the Fourier transform of a function that is 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), Schwartz (1966), Temple (1963), Gel'fand and Shilov (1964)). The regression function is then given by:⁹

$$g(x^*) = (2\pi)^{-1} \int \frac{\varepsilon_Y(\zeta)}{\phi(\zeta)} e^{-i\zeta x^*} d\zeta \quad (49)$$

where

$$\phi(\zeta) = \exp \left(\int_0^\zeta \frac{i\varepsilon_{(Z-X)Y,o}(\xi)}{\varepsilon_{Y,o}(\xi)} d\xi \right) \quad (50)$$

and where, for $V = Y$ or $(Z - X)Y$, the symbol $\varepsilon_{V,o}(\xi)$ 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 (44) and (45).

While there is some similarity between Equation (50) and Kotlarski's result (Equation (21)), there are important differences: First, (50) 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 (50), while only X and Z enter (21).

Schennach (2008) extends these results 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 analogues of Equations (42) and (43). 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 regular (rather than generalized) functions, so that the identification result can be directly used to deliver a nonparametric estimator. This estimator is shown

part of the generalized function (as shown, for instance, 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.

⁹The 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., Section 6 in Temple (1963). Note that the assumption that both $|E[Y | Z = z]|$ and $|g(x^*)|$ are bounded by polynomials implies that both $\varepsilon_Y(\zeta) = \gamma(\zeta)\phi(\zeta)$ 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.

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 and Hausman (1986)), but relatively little in nonlinear models, because the problem is not as straightforward. Hausman, Newey, Ichimura, and Powell (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:

$$Y_t = g(X_t^*, \theta) + \Delta Y_t \quad (51)$$

$$X_t = X_t^* + \Delta X_t \quad (52)$$

$$X_{t+1}^* = X_t^* + U_{t+1} \quad (53)$$

for $t = 1, \dots, T$, where $X_1^*, \Delta X_1, \dots, \Delta X_T, U_1, \dots, U_T, \Delta Y_1, \dots, \Delta Y_T$ 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^* :

$$X_t = X_t^* + \Delta X_t \quad (54)$$

$$X_{t+1} = X_{t+1}^* + \Delta X_{t+1} = X_t^* + (U_{t+1} + \Delta X_{t+1}) \quad (55)$$

where $X_t^*, \Delta X_t, (U_{t+1} + \Delta X_{t+1})$ are mutually independent, thus enabling the use of any Kotlarski-type identity. This can be extended by replacing (53) 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} , we have

$$X_t^* = X_{t+1} - (U_{t+1} + \Delta X_{t+1}) \quad (56)$$

$$X_t^* = X_{t-1} + (U_t - \Delta X_{t-1}) \quad (57)$$

in which none the error terms in parenthesis are independent from X_{t+1} or X_{t-1} . So the instrumental variable structure used in Hausman, Newey, Ichimura, and Powell (1991), Hausman, Newey, and Powell (1995), Newey (2001), Wang and Hsiao (2011), Schennach (2007a) and Schennach (2008) is not applicable. However, the more general approach of Hu and Schennach (2008) (discussed in Section 6.2 below) is applicable (for general generating processes for X_t^*), as noted by Wilhelm (2010), because it does *not* rely on the being able to write an instrument equation (of the form (36)) 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}^*) . A 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 upon a clever combination of operator inversion methods (Newey and Powell (2003), Darolles, Florens, and Renault (2002), Hall and Horowitz (2005), Carrasco, Florens, and Renault (2005)) with Fourier methods in quantile models (Schennach (2008)). The treatment of fully nonseparable panel data model in the presence of measurement error remains an open question.

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. We survey two possible approaches: One where stronger independence assumptions are made to secure identification, and one where 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 , ΔY are mutually independent, then it is well known that the standard regression model (2) 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), a large number of authors (e.g. Kapteyn and Wansbeek (1983), Kendall and Stuart (1979), Pal (1980), Cragg (1997), Lewbel (1997), Erickson and Whited (2002), Dagenais and Dagenais (1997), Erickson and Whited (2000), Lewbel (2007b), Bonhomme and Robin (2009), and the many references therein) have exploited this idea to identify and estimate a linear specification.

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

$$\theta = \frac{\text{Covar}[Z, Y]}{\text{Covar}[Z, X]} \quad (58)$$

where $Z = (X - E[X])(Y - E[Y])$. This expression has a natural instrumental variable interpretation, where 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 in Reiersol (1950) using c.f.

It should be noted that, 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 c.f., since the cumulants of a sum of independent random variables are simply the sum of their corresponding cumulants. Bonhomme and Robin (2009) considerably generalizes the treatment of Geary (1942) to multivariate factor models (introduced in Equation (35) above), enabling identification of the factor loadings with fewer measurements than from covariance information alone (Anderson and 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 (2) has only recently been solved (Schennach and Hu (2010)). 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 (i) the marginal c.f. of ΔX , ΔY , X^* , $g(X^*)$ do not vanish anywhere, (ii) $f_{X^*}(x^*)$ exists, (iii) $g(x^*)$ is continuously differentiable and (iv) $g'(x^*)$ 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

$$g(x^*) = a + b \ln(e^{cx^*} + d) \quad (59)$$

for some constants $a, b, c, d \in \mathbb{R}$, then the model is **identified**.

2. If $g(x^*)$ **is** of the form (59) with $d > 0$. Then the model is not identified if and only if X^* has a density of the form

$$f_{X^*}(x^*) = A \exp\left(-Be^{Cx^*} + CDx^*\right) \left(e^{Cx^*} + E\right)^{-F} \quad (60)$$

with $C \in \mathbb{R}$, $A, B, D, E, F \in [0, \infty)$ and ΔY has a type I extreme value factor¹⁰ (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 (59) 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 (59) 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 by Reiersol (1950)).

What is remarkable is that, among all possible (sufficiently regular) specifications, the nonidentified family is parametric with only 4 parameters (a, b, c, d in Equation (59)). 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.

¹⁰A 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 .

5.2 Bounding

Now, 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” 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), Manski (2003), Manski and Tamer (2002), Chernozhukov, Hong, and Tamer (2007)).

The bounds on the slope coefficient θ obtained in 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 $X^*, \Delta X, \Delta Y$, the full covariance structure of X and Y can be expressed in terms of the variances of the unobservable variables $\sigma_{X^*}^2, \sigma_{\Delta X}^2, \sigma_{\Delta Y}^2$:

$$\text{Var}[Y] = \theta^2 \sigma_{X^*}^2 + \sigma_{\Delta Y}^2 \quad (61)$$

$$\text{Var}[X] = \sigma_{X^*}^2 + \sigma_{\Delta X}^2 \quad (62)$$

$$\text{Covar}[X, Y] = \theta \sigma_{X^*}^2 \quad (63)$$

Combining (62) and (63) and using the constraint $\sigma_{\Delta X}^2 \geq 0$ yields (if $\theta \geq 0$)

$$\theta \geq \frac{\text{Covar}[X, Y]}{\text{Var}[X]} \quad (64)$$

while combining (61) and (63) and using the constraint $\sigma_{\Delta Y}^2 \geq 0$ yields (if $\theta \geq 0$)

$$\theta \leq \frac{\text{Var}[Y]}{\text{Covar}[X, Y]} \quad (65)$$

The directions on the inequalities in (64) and (65) are reversed if $\theta < 0$. These bounds have the straightforward interpretation as the slope coefficient of the “forward” regression of Y on X (Equation (64)) and of the reciprocal of slope coefficient of the “reverse” regression of X on Y (Equation (65)). These bounds are sharp under uncorrelatedness of $X^*, \Delta X, \Delta Y$.

This result has been elegantly generalized by Klepper and Leamer (1984) to multivariate linear regression where all regressors ($X_1^*, \dots, X_{d_X}^*$) are measured with error. It suffices to sequentially use each of the $d_{X^*} + 1$ observed variables (Y, X_1, \dots, X_{d_X}) as the dependent variable and remaining variables as regressors. After rearrangement, each regression line $k = 1, \dots, d_{X^*}$ 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 $\theta^{(1)}, \dots, \theta^{(d_{X^*})}$ 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*.¹¹ 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 and Leamer (1984), Erickson (1993)).

Unfortunately, the derivation of sharp bounds in 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 (2010) 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 and Henry (2006), Ekeland, Galichon, and Henry (2010) and Beresteanu, Molchanov, and Molinari (2010).

Schennach (2010) seeks to find the value(s) of a parameter vector θ that satisfy a set of moment conditions that are known to hold in the population:

$$E [m(U, Z, \theta)] = 0 \tag{66}$$

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 $Z = z$ while 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

¹¹ An orthant is a set of vectors whose elements share the same pattern of signs (it is the multivariate generalization of a quadrant).

loss of generality that \mathcal{U} does not depend on z , since this dependence could be incorporated into $m(u, z, \theta)$. Expectations are calculated under the distribution specified as a subscript. The key identify is as follows: For any θ and any distribution π ,

$$\inf_{\mu} \|E_{\mu \times \pi} [m(U, Z, \theta)]\| = 0 \quad (67)$$

if and only if

$$\inf_{\gamma \in \mathbb{R}^{d_m}} \|E_{\pi} [\tilde{m}(Z, \theta, \gamma)]\| = 0 \quad (68)$$

where

$$\tilde{m}(z, \theta, \gamma) \equiv \frac{\int m(u, z, \theta) \exp(\gamma' m(u, z, \theta)) d\rho(u|z)}{\int \exp(\gamma' m(u, z, \theta)) d\rho(u|z)} \quad (69)$$

where ρ is a user-specified conditional distribution such that (i) $\rho(\cdot|z)$ is supported on \mathcal{U} at each z and (ii) $E_{\pi} [\ln E_{\rho} [\exp(\gamma' m(U, Z, \theta)) | Z]]$ 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 by Proposition 1 in Schennach (2010).

This identity is useful, because the original problem of interest (67), which involves optimization over an infinite-dimensional quantity μ has been replaced by a finite dimensional optimization (68). This simplification is made possible by the fact that the so-called parametric least-favorable entropy maximizing family of distributions (proportional to $\exp(\gamma' m(u, z, \theta)) d\rho(u|z)$) used to compute the expectation over U in (69) 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 (since the optimization over γ would yield the same profiled objective function in terms of θ).

The intuition behind this result is that Equation (67) has potentially an infinite number of solutions. However, since we only need to find one, we can rank distributions according to some criterion (here, their entropy) and convert an “existence” problem into an optimization problem under the constraint (66). This constrained entropy maximization problem has a unique solution

which turns out to have the convenient form (69).

The simplest way to evaluate the integral (69) defining the moment function is to draw random vectors u_j , $j = 1, \dots, R$ from a density proportional to $\exp(\gamma' m(u, z, \theta)) d\rho(u|z)$ using, e.g., the Metropolis algorithm and calculate the average

$$\frac{1}{R} \sum_{j=1}^R m(u_j, z, \theta) \quad (70)$$

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 (69). 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 and 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 (66). 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., Newey and Smith (2004), Owen (1990), Imbens, Spady, and Johnson (1998), Kitamura and Stutzer (1997), Qin and Lawless (1994)) or Exponentially Tilted Empirical Likelihood (Schennach (2007b)). Existing generic inference techniques for set-identified models (such as Chernozhukov, Hong, and Tamer (2007)) 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 Figure 6 in Schennach (2010)) 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 being set-identified, as the data generating process changes. For instance, Figure 5 in Schennach (2010) shows that, in a linear regression with measurement error estimated using higher order moments, the objective function flattens over an interval (which is

indicative of set-identification¹²) when the distribution of the true regressor becomes normal, while 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 (censoring, truncation, interval-valued data, limited dependent variables, panel data with nonseparable correlated individual-specific heterogeneity, various game-theoretic models, etc.). Schennach (2010) 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, Brown, and Mathiowetz (2001), Bollinger (1998), Hyslop and 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 is, 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.

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 that 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), while Lewbel (2007a) considers treatment effect models when the treatment (a binary variable) is misclassified and an instrument is available.

¹²Although 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, Hong, and Tamer (2007)).

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 approach is a simpler finite-dimensional version of the approach described in Section 6.2 below). Gawade (2007) exploits a “three-way arrays” decomposition (Kruskal (1977)) to obtain a result similar to Hu (2008) that is more in the spirit of factor models, because all variables of the model are treated in a symmetric fashion.

The idea of using the full distribution of the variables to gain point identification without relying on auxiliary variables (as discussed in Section 5.1) has been used in regression models with discrete mismeasured regressors, as discussed in a series of papers culminating with Chen, Hu, and Lewbel (2009). However, the requisite identification conditions in Chen, Hu, and Lewbel (2009) are, in the authors’ words, “rather complicated”, as they involve checking the rank of a matrix constructed from products and sums of numerous matrices defined in terms of various functions of c.f.

It should be noted that identification results for discrete variables *do not* imply identification of the corresponding model with continuous variables via a limiting argument. To illustrate this, let $\text{Id}(M)$ denote a function equal “1” if model M is identified and “0” otherwise. This function is necessarily a discontinuous function of M because its range is discrete. Consequently, for a sequence of models M_n , we do not have $\lim_{n \rightarrow \infty} \text{Id}(M_n) = \text{Id}(\lim_{n \rightarrow \infty} M_n)$ in general. Therefore, when constructing a sequence of identified discrete models converging to a continuous model, we cannot conclude that the limiting continuous model is identified.

6.2 Continuously distributed variables

Hu and 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

$$f_{YX|Z}(y, x|z) = \int f_{Y|X^*}(y|x^*) f_{X|X^*}(x|x^*) f_{X^*|Z}(x^*|z) dx^* \quad (71)$$

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

$$f_{YXZ}(y, x, z) = \int f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*) dx^* \quad (72)$$

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, Heckman, and Schennach (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)).

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, Florens, and Renault (2005)), 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:

$$[L_{V|U}q](v) \equiv \int f_{V|U}(v|u) q(u) du, \quad (73)$$

The conditions needed for identification are (i) the random vectors X, Y, Z are mutually independent conditional on X^* , (ii) the operators $L_{X|X^*}$ and $L_{Z|X}$ are injective (iii) the distributions of Y given $X^* = x_1^*$ and of Y given $X^* = x_2^*$ differ whenever $x_1^* \neq x_2^*$. and (iv) there exists a known functional M such that $M[f_{X|X^*}(\cdot|x^*)] = x^*$. Assumption (i) is a fairly natural “exclusion” restriction. Assumption (ii) 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 and Powell (2003), Darolles, Florens, and Renault (2002), Hall and Horowitz (2005), Carrasco, Florens, and Renault (2005), Chernozhukov and Hansen (2005), Mattner (1993), d’Haultfoeuille (2006)). This assumption also (essentially) demands that the dimensions of X^* , X , and Z be the same. Assumption (iii) is weaker than injectivity and can be satisfied even if Y is scalar. Remarkably, Assumptions (ii) and (iii) 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 and Rubin (1956)). Assumption

(iv) generalizes classical measurement error (corresponding to “ M ” being the mean) to other types of centering restrictions (such as the mode, the median or any other quantile). As discussed in Hu and Schennach (2008), median and mode restrictions are well supported by existing validation data evidence (Bound and Krueger (1991)). In addition, such assumptions are robust to other data problems: For instance, under weak conditions, the mode is unaffected by truncation, while the median is unaffected by censoring.

The proof of identification in Hu and Schennach (2008) can be outlined as follows. Assumption (i) directly implies the integral Equation (71). This equation can be cast as an operator equivalence relationship:

$$L_{y;X|Z} = L_{X|X^*} D_{y;X^*} L_{X^*|Z}, \quad (74)$$

where $L_{y;X|Z}$ is defined analogously to $L_{X|Z}$ with $f_{X|Z}$ replaced by $f_{Y,X|Z}(y, \cdot)$ for a given y and where $D_{y;X^*}$ 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 (since $f_{X|Z}(x|z) = \int f_{X|X^*}(x|x^*)f_{X^*|Z}(x^*|z)dx^*$, again by conditional independence). Isolating $L_{X^*|Z}$ to yield

$$L_{X^*|Z} = L_{X|X^*}^{-1} L_{X|Z}, \quad (75)$$

substituting it into (74) and rearranging, we obtain:

$$L_{y;X|Z} L_{X|Z}^{-1} = L_{X|X^*} D_{y;X^*} L_{X|X^*}^{-1}, \quad (76)$$

where all inverses can be shown to exist over suitable domains¹³ under the injectivity assumptions made.

Equation (76) states that the operator $L_{Y;X|Z} L_{X|Z}^{-1}$ admits a *spectral decomposition* (eigenvalue-eigenfunction decomposition). The operator to be diagonalized is defined in terms of observable densities, while 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 and Schennach (2008) employ four techniques. First, a powerful result from spectral analysis (Theorem XV 4.5 in Dunford and Schwartz (1971)) 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

¹³Injectivity of $L_{Z|X}$ is related to existence of $L_{X|Z}^{-1}$ — see Lemma 1 in Hu and Schennach (2008).

degenerate eigenvalues are present, one can use Assumption (iii) 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, in order to uniquely determine the ordering and indexing of the eigenvalues and eigenfunctions, one can invoke Assumption (iv): Consider another variable \tilde{x}^* related to x^* through $x^* = R(\tilde{x}^*)$, and note that

$$M \left[f_{X|\tilde{X}^*}(\cdot|\tilde{x}^*) \right] = M \left[f_{X|X^*}(\cdot|R(\tilde{x}^*)) \right] = R(\tilde{x}^*), \quad (77)$$

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 $M \left[f_{X|X^*}(\cdot|x^*) \right]$ strictly increasing in x^* , because, in finite dimension, ordering the eigenvectors is sufficient, while in infinite dimensions, it is possible to reparametrize 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 (75) implies that $f_{X^*|Z}(x^*|z)$ is also identified. Since the identities (76) and (75) use and provide the same information as Equation (71), this establishes uniqueness of the solution to Equation (71). Equation (72) follows by similar manipulations. Hu and Schennach (2008) also suggest an estimator based on Equation (71) or (72) obtained by substituting series approximations for the unknown densities.

It may be useful to note that operator diagonalization techniques generalize Fourier transforms, which only diagonalize convolution operators (whose eigenfunctions are always known since 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 (Cunha, Heckman, and Schennach (2010), Hu and Shum (2010)) and Berkson-type errors (Schennach (2009)), discussed below.

6.3 Berkson measurement error

Berkson (1950) introduced 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 of the form

$$X^* = X + \Delta X^* \quad (78)$$

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 and Imbens (2001), Hoderlein and Winter (2010)). Another plausible occurrence is when, for instance, only a regional average is observed for one of the regressors, although the researcher is interested in a model at the individual level. It has also been suggested that some measurement errors in the regressors can consist of a combination of a classical and Berkson error (see, for instance, Hyslop and Imbens (2001)).

It is known 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 (Buonaccorsi and Lin (2002)). 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 and Huang (2000)) and, more recently, for a very wide range of parametric models (Wang (2004), Wang (2007)). Nonparametric estimation, but under the assumption that the distribution of the Berkson error ΔX^* is fully known, has also been demonstrated (Delaigle, Hall, and Qiu (2006), Carroll, Delaigle, and Hall (2007)).

Schennach (2009) 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 that is related to the true regressor X^* via

$$Z = h(X^*) + \Delta Z \tag{79}$$

where the function $h(\cdot)$ is also unknown. The following assumptions are made: (i) X , ΔX^* , ΔY , ΔZ are mutually independent (this can be relaxed, as explained in the paper), (ii) ΔX^* , ΔY , ΔZ are centered (i.e. have zero mean, mode, median, etc.), (iii) the marginal c.f. of ΔX^* and ΔZ are nonvanishing, (iv) the functions $g(x^*)$ and $h(x^*)$ are one-to-one, (v) h and its inverse are differentiable. Identification can then be stated as: Given the true observed conditional density

$f_{Y,Z|X}$, the functional equation

$$f_{Y,Z|X}(y, z|x) = \int f_{\Delta Z}(z - h(x^*)) f_{\Delta Y}(y - g(x^*)) f_{\Delta X^*}(x^* - x) dx^* \quad (80)$$

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

This result is proven using operator techniques analogous to Hu and Schennach (2008), but with some crucial differences. First, following steps similar to Section 6.2, Equation (80) implies the operator equivalence

$$L_{y;Z|X} L_{Z|X}^{-1} = L_{Z|X^*} D_{y;X^*} L_{Z|X^*}^{-1} \quad (81)$$

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

$$L_{X^*|X} = L_{Z|X^*}^{-1} L_{Z|X} \quad (82)$$

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

6.4 Bounding

If the conditions needed for point identification are too stringent, another way to approach non-classical 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), Manski (2003), Chernozhukov, Hong, and Tamer (2007), Imbens and Manski (2004)). For instance, respondents in a survey may report ranges instead of specific values (i.e., interval-valued data) 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 and Tamer (2002), who derive bounds on the regression coefficients when both the regressors and the dependent variable could be interval-valued. Also, Horowitz and 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 char-

acterizes the identified set under a wide range of plausible *a priori* assumptions regarding the measurement error (such as bounds on the misclassification probabilities, symmetry, monotonicity constraints, etc.).

The ELVIS method (Schennach (2010)), 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 narrowing down the identified set, without having to revisit the entire analysis of identification. Figures 2 and 3 in Schennach (2010) illustrates this idea for interval-valued data and censored regressions, where higher order moments constraints (which impose plausible constraints on the form of heteroskedasticity) are shown to 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. The usefulness of ELVIS is not limited to higher order moments — it can just as easily handle the median restrictions on the measurement error suggested in Hu and 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, Perrigne, and Vuong (2000) Athey and Haile (2002) Krasnokutskaya (2011) An, Hu, and Shum (2010) Hu, McAdams, and Shum (2010)). 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 deliver identification of the distribution of the random coefficients and time-varying errors distributions (Arellano and Bonhomme (2011)). Treatment effect models where the control variables only enable imperfect matching can also benefit from measurement error techniques (Heckman, Schennach, and Williams (2010)). 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 (2010)), such as interval-valued

data, limited dependent variables, panel data and some game-theoretic models.

8 Conclusion

This paper has overviewed 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 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 (2010)) 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 datasets (with potential repeated measurements or instruments) are becoming more common.

Acknowledgements

The author acknowledges support from NSF grants SES-0752699 and SES-1061263.

References

- AMEMIYA, Y. (1985): “Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model,” *Journal of Econometrics*, 28, 273–289.
- AN, Y., Y. HU, AND M. SHUM (2010): “Estimating First-Price Auctions with an Unknown Number of Bidders: A Misclassification Approach,” *Journal of Econometrics*, 157, 328–341.
- ANDERSON, T. W., AND H. RUBIN (1956): “Statistical inference in factor analysis,” *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*, V, 111–150.
- ARELLANO, M., AND S. BONHOMME (2011): “Identifying Distributional Characteristics in Random Coefficients Panel Data Models,” *Review of Economic Studies*, 79, 987–1020.
- ATHEY, S., AND P. A. HAILE (2002): “Identification of Standard Auction Models,” *Econometrica*, 70, 2107–2140.
- BATTISTIN, E., AND A. CHESHER (2009): “Treatment Effect Estimation with Covariate Measurement Error,” Working Paper CWP25/09, Cemmap.
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2010): “Sharp identification regions in models with convex moment predictions,” Working Paper CWP25/10, Cemmap.

- BERKSON, J. (1950): "Are there two regressions?," *Journal of the American Statistical Association*, 45, 164–180.
- BOLLINGER, C. R. (1998): "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics*, 16, 576–594.
- BONHOMME, S., AND J.-M. ROBIN (2009): "Consistent Noisy Independent Component Analysis," *Journal of Econometrics*, 149, 12–25.
- (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *Review of Economic Studies*, 77, 491–533.
- BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. V. Elsevier Science.
- BOUND, J., AND A. B. KRUEGER (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right," *Journal of Labor Economics*, 9, 1–24.
- BROWNING, M., AND T. F. CROSSLEY (2009): "Are two cheap, noisy measures better than one expensive, accurate one?," Working Paper W09/01, Institute for Fiscal Studies.
- BUONACCORSI, J. P., AND C.-D. LIN (2002): "Berkson measurement error in designed repeated measures studies with random coefficients," *Journal of Statistical Planning and Inference*, 104, 53–72.
- CARRASCO, M., J.-P. FLORENS, AND E. RENAULT (2005): "Linear Inverse Problems and Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, vol. Vol. 6. Elsevier Science.
- CARROLL, R., AND P. HALL (1988): "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83(404), 1184–1186.
- CARROLL, R. J., A. DELAIGLE, AND P. HALL (2007): "Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors," *Journal of the Royal Statistical Society B*, 69, 859–878.
- CHEN, X., H. HONG, AND E. TAMER (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343–366.
- CHEN, X., Y. HU, AND A. LEWBEL (2009): "Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information," *Statistica Sinica*, 19, 949–968.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets In Econometric Models," *Econometrica*, 75, 1243–1284.
- CHESHER, A. (1991): "The Effect of Measurement Error," *Biometrika*, 78, 451.
- (2001): "Parameter Approximations for Quantile Regressions with Measurement Error," Working Paper CWP02/01, Department of Economics, University College London.
- CRAGG, J. C. (1997): "Using higher moments to estimate the simple errors-in-variables model," *Rand Journal of Economics*, 28, S71–S91.
- CUNHA, F., J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931f.
- DAGENAIS, M. G., AND D. L. DAGENAIS (1997): "Higher Moment Estimators for Linear Regression Models with Errors in Variables," *Journal of Econometrics*, 76, 193–221.
- DAROLLES, S., J.-P. FLORENS, AND E. RENAULT (2002): "Nonparametric Instrumental Regres-

- sion,” Working Paper 05-2002, Centre de Recherche et Développement en Économique.
- DELAIGLE, A., P. HALL, AND A. MEISTER (2008): “On Deconvolution with Repeated Measurements,” *Annals of Statistics*, 36, 665–685.
- DELAIGLE, A., P. HALL, AND P. QIU (2006): “Nonparametric methods for solving the Berkson errors-in-variables problem,” *Journal of the Royal Statistical Society B*, 68, 201–220.
- D’HAULTFOEUILLE, X. (2006): “On the Completeness Condition in Nonparametric Instrumental Problems,” Working Paper, ENSAE, CREST-INSEE and Université de Paris I.
- DUNFORD, N., AND J. T. SCHWARTZ (1971): *Linear Operators*. John Wiley & Sons, NY.
- EKELAND, I., A. GALICHON, AND M. HENRY (2010): “Optimal transportation and the falsifiability of incompletely specified economic models,” *Economic Theory*, 42, 355–374.
- ERICKSON, T. (1993): “Restricting regression slopes in the errors-in-variables model by bounding the error correlation,” *Econometrica*, 61, 959–969.
- ERICKSON, T., AND T. M. WHITED (2000): “Measurement Error and the Relationship between Investment and “q,”” *Journal of Political Economy*, 108, 1027–1057.
- (2002): “Two-step GMM estimation of the errors-in-variables model using high-order moments,” *Econometric Theory*, 18, 776–799.
- EVDOKIMOV, K. (2009): “Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity,” Working Paper, Yale University.
- EVDOKIMOV, K., AND H. WHITE (2011): “An Extension of a Lemma of Kotlarski,” *Econometric Theory*, accepted.
- FAN, J. (1991a): “Asymptotic Normality for Deconvolution Kernel Density Estimators,” *Sankhyā: The Indian Journal of Statistics*, 53, Series A, Pt. 1, 97–110.
- (1991b): “On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems,” *Annals of Statistics*, 19(3), 1257–1272.
- FAN, J., AND Y. K. TRUONG (1993): “Nonparametric Regression with Errors in Variables,” *Annals of Statistics*, 21(4), 1900–1925.
- FRISCH, R. (1934): *Statistical Confluence Analysis by Means of Complete Regression Systems*. University Institute of Economics, Oslo, Norway.
- GALICHON, A., AND M. HENRY (2006): “Dilation Bootstrap: A methodology for constructing confidence regions with partially identified models,” Working Paper, Harvard University and Columbia University.
- GAWADE, N. G. (2007): “Measurement Error in Discrete Explanatory Variables: Implications of Conditional Independence,” Working Paper, Princeton University.
- GEARY, R. C. (1942): “Inherent Relations Between Random Variables,” *Proceedings of the Royal Irish Academy*, 47A, 63–76.
- GEL’FAND, I. M., AND G. E. SHILOV (1964): *Generalized Functions*. Academic Press, New York.
- GRILICHES, Z., AND J. A. HAUSMAN (1986): “Errors in Variables in panel data,” *Journal of Econometrics*, 31, 93–118.
- GRILICHES, Z., AND V. RINGSTAD (1970): “Errors-in-Variables Bias in Nonlinear Contexts,” *Econometrica*, 38, 368–370.
- HALL, P., AND J. L. HOROWITZ (2005): “Nonparametric Methods for Inference in the Presence of Instrumental Variables,” *Annals of Statistics*, 33, 2904–2929.
- HANSEN, L. P. (1982): “Large sample properties of generalized method of moment estimators,” *Econometrica*, 50, 1029–1054.

- HAUSMAN, J. (2001): “Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left,” *Journal of Economic Perspectives*, 15(4), 57–67.
- HAUSMAN, J., W. NEWEY, H. ICHIMURA, AND J. POWELL (1991): “Measurement Errors in Polynomial Regression Models,” *Journal of Econometrics*, 50, 273–295.
- HAUSMAN, J., W. NEWEY, AND J. POWELL (1995): “Nonlinear Errors in Variables. Estimation of Some Engel Curves,” *Journal of Econometrics*, 65, 205–233.
- HECKMAN, J. J., S. M. SCHENNACH, AND B. D. WILLIAMS (2010): “Matching on proxy variables,” Working Paper, University of Chicago.
- HODERLEIN, S., AND J. WINTER (2010): “Structural Measurement Errors in Nonseparable Models,” *Journal of Econometrics*, 157, 432–440.
- HOROWITZ, J., AND C. MANSKI (1995): “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica*, 63, 281–302.
- HOROWITZ, J. L., AND M. MARKATOU (1996): “Semiparametric Estimation of Regression Models for Panel Data,” *Review of Economic Studies*, 63, 145.
- HSIAO, C. (1989): “Consistent Estimation for Some Nonlinear Errors-in-Variables Models,” *Journal of Econometrics*, 41, 159–185.
- HU, Y. (2008): “Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution,” *Journal of Econometrics*, 144, 27–61.
- HU, Y., D. MCADAMS, AND M. SHUM (2010): “Nonparametric identification of first-price auctions with non-separable unobserved heterogeneity,” Working Paper 553, Johns Hopkins University.
- HU, Y., AND G. RIDDER (2010): “On Deconvolution as a First Stage Nonparametric Estimator,” *Econometric Reviews*, 29, 1–32.
- (2012): “Estimation of Nonlinear Models with Measurement Error Using Marginal Information,” *Journal of Applied Econometrics*, 27, 347–85.
- HU, Y., AND S. M. SCHENNACH (2008): “Instrumental Variable Treatment of Nonclassical Measurement Error Models,” *Econometrica*, 76, 195–216.
- HU, Y., AND M. SHUM (2010): “Nonparametric identification of dynamic models with unobserved state variables,” Working Paper, Johns Hopkins University.
- HUWANG, L., AND Y. H. S. HUANG (2000): “On errors-in-variables in polynomial regressions — Berkson case,” *Statistica Sinica*, 10, 923–936.
- HYSLOP, D. R., AND G. W. IMBENS (2001): “Bias from classical and other forms of measurement error,” *Journal of Business & Economic Statistics*, 19, 475–481.
- IMBENS, G., AND C. MANSKI (2004): “Confidence intervals for partially identified parameters,” *Econometrica*, 72, 1845–1857.
- IMBENS, G. W., R. H. SPADY, AND P. JOHNSON (1998): “Information Theoretic Approaches to Inference in Moment Condition Models,” *Econometrica*, 66, 333–357.
- KAPTEYN, A., AND T. WANSBEEK (1983): “Identification in the Linear Errors in Variables Model,” *Econometrica*, 51, 1847–1849.
- KENDALL, M. G., AND A. STUART (1979): *The Advanced Theory of Statistics*. Macmillan, New York, 4th edition edn.
- KITAMURA, Y., AND M. STUTZER (1997): “An Information-Theoretic Alternative to Generalized Method of Moment Estimation,” *Econometrica*, 65, 861–874.
- KLEPPER, S., AND E. E. LEAMER (1984): “Consistent Sets of Estimates for Regressions with Errors in all Variables,” *Econometrica*, 52, 163–183.

- KOTLARSKI, I. (1967): “On Characterizing The Gamma And The Normal Distribution,” *Pacific Journal Of Mathematics*, 20, 69–76.
- KRASNOKUTSKAYA, E. (2011): “Identification and Estimation in Procurement Auctions under Unobserved Auction Heterogeneity,” *Review of Economic Studies*, 78, 293–327.
- KRUSKAL, J. B. (1977): “Three-Way Arrays: Rank and Uniqueness of Trilinear Decompositions, with Applications to Arithmetic Complexity and Statistics,” *Linear Algebra and its Applications*, 18, 95–138.
- LACZKOVICH, M. (1984): “Differentiable restrictions of Continuous Functions,” *Acta Mathematica Hungarica*, 44, 355.
- 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(5), 1201–1213.
- (2007a): “Estimation of Average Treatment Effects With Misclassification,” *Econometrica*, 75, 537–551.
- (2007b): “Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models,” Working Paper, Boston College.
- LI, T. (2002): “Robust and consistent estimation of nonlinear errors-in-variables models,” *Journal of Econometrics*, 110, 1–26.
- LI, T., I. PERRIGNE, AND Q. VUONG (2000): “Conditionally independent private information in OCS wildcat auctions,” *Journal of Econometrics*, 98, 129.
- LI, T., AND Q. VUONG (1998): “Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators,” *Journal of Multivariate Analysis*, 65, 139–165.
- LIGHTHILL, M. J. (1962): *Introduction to Fourier Analysis and Generalized Function*. London: Cambridge University Press.
- LIU, M., AND R. TAYLOR (1989): “A consistent nonparametric density estimator for the deconvolution problem,” *The Canadian Journal of Statistics*, 17(4), 427–438.
- LOÈVE, M. (1977): *Probability Theory I*. New York: Springer.
- LUKACS, E. (1970): *Characteristic Functions*. Griffin, London, second edn.
- MAHAJAN, A. (2006): “Identification and Estimation of Single Index Models with Misclassified Regressor,” *Econometrica*, 74, 631–665.
- MANSKI, C. (1990): “Nonparametric bounds on Treatment Effects,” *American Economic Review Papers and Proceedings*, 80, 319–323.
- (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.
- MANSKI, C. F., AND E. TAMER (2002): “Inference on Regressions with Interval data on a Regressor or Outcome,” *Econometrica*, 70, 519–546.
- MATTNER, L. (1993): “Some incomplete but boundedly complete location families,” *Annals of Statistics*, 21, 2158–2162.
- MATZKIN, R. L. (2003): “Nonparametric Estimation of Nonparametric Nonadditive Random Functions,” *Econometrica*, 71, 1339–1375.
- MCFADDEN, D. (1989): “A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration,” *Econometrica*, 57, 995–1026.
- MCINTYRE, J., AND L. A. STEFANSKI (2011): “Heteroscedastic measurement deconvolution,” *Annals of the Institute of Statistical Mathematics*, 63, 81–99.
- MOLINARI, F. (2008): “Partial Identification of Probability Distributions with Misclassified Data,”

- Journal of Econometrics*, 144, 81–117.
- NEWHEY, W. (2001): “Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models,” *Review of Economics and Statistics*, 83, 616–627.
- NEWHEY, W., AND R. J. SMITH (2004): “Higher-Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, 72, 219–255.
- NEWHEY, W. K., AND J. L. POWELL (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71, 1565–1578.
- OWEN, A. B. (1988): “Empirical Likelihood Ratio Confidence Intervals for a Single Functional,” *Biometrika*, 75, 237–249.
- (1990): “Empirical Likelihood Ratio Confidence Regions,” *Annals of Statistics*, 18, 90–120.
- PAKES, A., AND D. POLLARD (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027–1057.
- PAL, M. (1980): “Consistent moment estimators of regression-coefficients in the presence of errors in variables,” *Journal of Econometrics*, 14, 349–364.
- QIN, J., AND J. LAWLESS (1994): “Empirical Likelihood and General Estimating Equations,” *Annals of Statistics*, 22, 300–325.
- RAO, P. (1992): *Identifiability in Stochastic Models*. New York: Academic Press.
- REIERSOL, O. (1950): “Identifiability of a Linear Relation between Variables Which Are Subject to Error,” *Econometrica*, 18, 375–389.
- RIDDER, G., AND R. MOFFITT (2006): “The econometrics of data combination,” in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6B. Elsevier Science.
- SCHENNACH, S. M. (2000): “Estimation of Nonlinear Models with Measurement Error,” Working Paper, University of Chicago.
- (2004a): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72, 33–75.
- (2004b): “Exponential specifications and measurement error,” *Economics Letters*, 85, 85–91.
- (2004c): “Nonparametric Estimation in the Presence of Measurement Error,” *Econometric Theory*, 20, 1046–1093.
- (2007a): “Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models,” *Econometrica*, 75, 201–239.
- (2007b): “Point Estimation with Exponentially Tilted Empirical Likelihood,” *Annals of Statistics*, 35, 634–672.
- (2008): “Quantile regression with mismeasured covariates,” *Econometric Theory*, 24, 1010–1043.
- (2009): “Instrumental variable treatment of the nonparametric Berkson measurement error model,” Working Paper, University of Chicago.
- (2010): “Entropic Latent Variable Integration via Simulation,” Working Paper, University of Chicago.
- SCHENNACH, S. M., AND Y. HU (2010): “Identification and estimation of a nonparametric relation between variables which are subject to errors,” Working Paper, University of Chicago.
- SCHENNACH, S. M., H. WHITE, AND K. CHALAK (2012): “Local indirect least squares and average marginal effects in nonseparable structural systems,” *Journal of Econometrics*, pp. 282–302, 166.
- SCHENNACH, S. M., AND B. D. WILLIAMS (2010): “Nonparametric Factor Score Regression with

- an Application to the Technology of Skill Formation,” Working Paper, University of Chicago.
- SCHWARTZ, L. (1966): *Théorie des distributions*. Paris, Hermann.
- SEPANSKI, J. H., AND R. J. CARROLL (1993): “Semiparametric Quasi-likelihood and Variance Function Estimation in Measurement Error Models,” *Journal of Econometrics*, 58, 223–256.
- TAUPIN, M.-L. (2001): “Semi-Parametric Estimation in the Nonlinear Structural Errors-in-Variables Model,” *Annals of Statistics*, 29, 66–93.
- TEMPLE, G. (1963): “The Theory of Weak Functions. I,” *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 276, 149–167.
- WANG, L. (2004): “Estimation of nonlinear models with Berkson measurement errors.,” *Annals of Statistics*, 32, 2559–2579.
- (2007): “A unified approach to estimation of nonlinear mixed effects and Berkson measurement error models,” *Canadian Journal of Statistics*, 35, 233–248.
- WANG, L., AND C. HSIAO (2011): “Method of moments estimation and identifiability of nonlinear semiparametric errors-in-variables models,” *Journal of Econometrics*, 165, 30–44.
- WILHELM, D. (2010): “Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error,” Working Paper, University of Chicago.