cemmap
centre for microdata methods and practice

Measurement systems

Susanne Schennach

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP12/21

UKRI Economic and Social Research Council

Measurement systems

Susanne Schennach*
Brown University

March 6, 2021

Abstract

Economic models often depend on quantities that are unobservable, either for privacy reasons or because they are difficult to measure. Examples of such variables include human capital (or ability), personal income, unobserved heterogeneity (such as consumer "types"), etc. This situation has historically been handled either by simply using observable imperfect proxies for each the unobservables, or by assuming that such unobservables satisfy convenient conditional mean or independence assumptions that enable their elimination from the estimation problem. However, thanks to tremendous increases in both the amount of data available and computing power, it has become possible to take full advantage of recent formal methods to infer the statistical properties of unobservable variables from multiple imperfect measurements of them.

The general framework used is the concept of *measurement systems* in which a vector of observed variables is expressed as a (possibly nonlinear or nonparametric) function of a vector of all unobserved variables (including unobserved error terms or "disturbances" that may have non additively separable affects). The framework emphasizes important connections with related fields, such as nonlinear panel data, limited dependent variables, game theoretic models, dynamic models and set-identification. This review reports the progress made towards the central question of whether there exist plausible assumptions under which one can identify the joint distribution of the unobservables from the knowledge of the joint distribution of the observables. It also overviews empirical efforts aimed at exploiting such identification results to deliver novel findings that formally account for the unavoidable presence of unobservables.

*This work is supported by the US National Science Foundation under grants SES-1659334 and SES-1950969. The author would like to thank Vincent Starck for helpful comments.

1 Introduction

Economists have long understood that economic behavior is largely determined by quantities that are difficult to measure accurately or are entirely unobserved (e.g., Griliches and Ringstad (1970), Amemiya (1985), Bound, Brown, and Mathiowetz (2001), Hausman (2001), Aigner, Hsiao, Kapteyn, and Wansbeek (1986)). Fortunately, techniques to handle such situations have been under constant development for a long time and, in fact, have experienced a recent surge in interest in empirical applications, thanks in part to the availability of rich datasets, increased computing power and ongoing theoretical advances. Recognizing the presence of unobservables enables a better assessment of the true uncertainty associated with economic modeling and paves the way for more informed policy decisions and more representative counterfactual analyses.

Although other surveys on measurement error or latent variables exist (Carroll, Ruppert, Stefanski, and Crainiceanu (2006), Fuller (1987), Cheng and Ness (1999), Söderström (2018), Bound, Brown, and Mathiowetz (2001), Chen, Hong, and Nekipelov (2011), Wansbeek and Meijer (2000), Hu (2015), Schennach (2013a), Schennach (2016), Schennach (2018)), this review has a distinct focus. It emphasizes methods, through both theory and empirical examples, that target empirically relevant situations where the observed and unobserved variables can be multidimensional and where the true values of some variables are simply not accessible to the researchers. A natural framework to discuss these techniques is the concept of *measurement system*, in which information regarding the unobserved variables can be inferred from numerous observable variables. The motivation for focusing on measurement systems is twofold. First, in the age of "big data", high-dimensional data is becoming increasingly available, thus providing a rich source of multivariate data to choose from. Second, multidimensional settings offer more opportunity to uncover information about unobservable variables, because even a small signal contained in multiple measurements can suffice to recover information about the unobservable variables by exploiting the fact that the "signal" is common to multiple measurements while the "noise" is not. Measurement systems provide a unifying concept that highlights the connection between measurement error models and unobserved (or latent) variable models and emphasizes natural connections with other topics such as factor models, panel data, limited dependent variables and set-identification.

Special attention will be devoted to techniques that can readily be used in multivariate settings (either directly or through natural extensions introduced herein) and that avoid the traditional assumptions of classical zero-mean errors, linearity of the

model and/or additive separability of the errors. The main focus will be on identification issues, because these typically constitute the first hurdle in even defining a useful model of data generating process and also suggest novel ways to collect data that elicit many, though perhaps noisy, measures rather than a single but elusive accurate measure. As it will become clear, the minimum number of measurements strictly needed for identification is often not very large, and the redundancy provided by more numerous measures can be exploited to further improve statistical accuracy.

In this paper, we first characterize the various types of measurements and measurement systems that have been considered in the literature. We then describe the techniques available to establish that these measurement systems can be used to identify features of the underlying unobserved variables, using well-known applications to illustrate how the necessary information can typically be obtained. These approaches are primarily organized by the type of information they rely on, rather than by mathematical techniques they use. Finally, a more in-depth discussion of some these methods is given through a number of applications that demonstrate the effectiveness of measurement systems in addressing questions of empirical relevance.

2 Key concepts

2.1 Measurement systems

The overarching theme in this review is the concept of measurement system:

$$X = m\left(X^*, \Delta X\right) \tag{1}$$

which is represented by a (possibly nonlinear) function $m\left(\cdot,\cdot\right)$ that relates the observed data vector X to the unobserved data vectors X^* and ΔX. The vector X^* comprises all the variables of interest while the vector ΔX contains all the sources of noise. The mapping $m\left(\cdot,\cdot\right)$ may, in general, depend on some other observable vector C. However, for the purpose of studying identification, one can formally eliminate the vector C by working with conditional quantities throughout with the understanding that all quantities have an implicit C-dependence. (It may sometime be more convenient to express a measurement system by instead isolating some of the unobserved variables on the "left-hand side" in Equation (1).)

A distinguishing feature of a measurement system is that the support of the observables X is smaller (for instance in terms of dimensionality or cardinality) than the support of all the unobservables $X^*, \Delta X$. As a result, one cannot simply "invert" m to recover the unobservables and the system is fundamentally underdetermined. This

feature also makes measurement systems starkly distinct from well-known simultaneous equation models, where the unobserved disturbances can typically be recovered from the observables (for a given value of the model's parameters and the observable variables and under suitable monotonicity assumptions). To make identification matters even more challenging, the form of the function $m(\cdot,\cdot)$ may or may not be known a priori, although some restrictions on its behavior can be a priori placed, based on either economic theory or plausible normalizations.

In such a measurement system, one cannot hope to assign a specific value of X^* to each observation unit of the sample, but one may be able to make inference regarding its statistical properties, such as its mean, variance, quantiles, distribution, etc., or bounds on such quantities, potentially after conditioning on other observed or unobserved variables.

A central question in the study of measurement systems is thus one of identifiability: Under which constraints on the function m and on the joint distribution of the observables and unobservables can one recover features of interest of the distribution of X^*? The array of possible assumptions on m and on the distributions lead to the different types of measurements and models that have been considered in the literature and that will be discussed next.

2.2 Notation and conventions

We first briefly introduce the notations that will be used throughout. 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 density function and the distribution function of the random variable V and similarly for random vectors, whose dimension is denoted d_V. A similar notation is used for conditional quantities, e.g. $f_{V_1|V_2}(v_1|v_2)$. Although not strictly necessary, we will present most of the ideas assuming that all variables admit a density with respect to the Lebesgue measure. We disregard "almost everywhere" qualifications to lighten the text and avoid detailed definitions of suitable function spaces. We let $\mathbf{1}\{A\}$ denote an indicator function equal to 1 if A is true and equal to 0 otherwise while $E[\cdot]$ is the usual expectation. Norms are denoted by $\|\cdot\|$ and refer to the square root of the sum over all elements of a vector or a matrix. Abbreviations are consistently used for Generalized Method of Moment (GMM) and cumulative probability density function (cdf).

2.3 Classical measurement error

In classical measurement error, the unobserved but true value of the random variable of interest X^* and its observed but mismeasured counterpart X are related through

$$X = X^* + \Delta X \tag{2}$$

where ΔX denotes the measurement error. (Each of these random variables could, in general, be random vectors.) In this classical case, the measurement function $m(X^*, \Delta X)$ is linear and additively separable in its arguments. Two alternative types of *classical measurement error* have been described in the literature (here we adopt the terminology proposed in Schennach (2016)):

Definition 1 *In Equation (2), the measurement X (or the error ΔX) is said to be*

1. *strongly classical* if ΔX is independent from X^* and $E[\Delta X] = 0$ and
2. *weakly classical* if $E[\Delta X | X^* = x^*] = 0$ for all x^*

Both conditions imply the standard uncorrelatedness between ΔX and X^* that is typically assumed in *linear* measurement error models. Depending on the specific model, these two assumptions may include independence from, or conditioning on, some of the other variables of the model. The term *differential error* is sometimes used to describe errors that, albeit classical, are correlated with variables other than X^*. In the absence of such correlations, errors would be called *nondifferential*. When X^* and X are discrete, the term *misclassification* is often used in place of measurement error. While most of this review targets the more challenging case of continuous variables, some misclassification problems will also be considered.

Although in a measurement system, the dimension of all unobservables $(X^*, \Delta X)$ exceed that of X, the most fruitful cases (in terms of being able to make inferences about X^*) occur when the dimension of X is larger than the dimension of X^* alone (the unobserved variables of interest), as this may enable disentangling the signal X^* from the noise ΔX. To emphasize this, we shall express the left-hand-side vector as blocks, e.g., $(X', Z')'$ where X has the dimension of X^* and Z contains additional observable variables. This natural notation ties in with the concepts of repeated measurements, which are arguably one of the most effective ways to handle measurement error models (Schennach (2004a)).

2.4 Nonclassical measurement error

Errors that do not conform to the classical assumptions of the previous section are labelled *nonclassical measurement error*, which have long been a concern (Bound, Brown, and Mathiowetz (2001), Bollinger (1998), Bound and Krueger (1991)), but have only recently started to be addressed via formal methods. Nonclassical errors occur whenever the errors ΔX exhibit a bias that depends on the true X^* ($E[\Delta X|X^*] \neq E[\Delta X]$). A well-known empirical example is self-reported income, where respondents tend to over report at the lower end of the income distribution and under report at the upper end (Bollinger (1998)). A plausible model for the generation of variables contaminated by nonclassical errors is the presence of a random noise that does not affect the observed outcome in an additive fashion, leading to a general nonlinear nonseparable model (Matzkin (2003), Chesher (2003), Matzkin (2008), Heckman and Vytlacil (2005), Chernozhukov, Imbens, and Newey (2007)), which also forms a measurement system:

$$X = m(X^*, \Delta X) \tag{3}$$

where $m(\cdot, \cdot)$ is a general nonlinear function. In this more general setting, one obviously needs some form of normalization, for otherwise, one can always change the function $m(\cdot, \cdot)$ and the distribution of the unobservables in ways that exactly offset each other. For instance, one could select a normalization so that ΔX is independent from X^* and let $m(\cdot, \cdot)$ account for the dependence.

Measurement with nonclassical errors include *indicators (or proxies)* which are related to the true value of the variable of interest but may be expressed in different units or may even be nonlinearly related to the true value. Indicators are often monotone in the unobserved variable they proxy for. In a *factor model*, the indicators may proxy for multiple unobserved variables — i.e., one may not know in advance which indicator provides information regarding which unobserved variable.

The special important case of linear systems or factor models (i.e. $X = \Lambda X^* + \Delta X$ for some matrix Λ of "factor loadings") has a long history (Spearman (1904), Anderson and Rubin (1956)). Such class of systems includes measurements that are not fully classical, because the loadings may differ from unity ($E[X - X^*|X^*] = (\Lambda - I) X^* \neq 0$ if $\Lambda \neq I$). Extensions of this classic linear setting to general nonlinear and even nonseparable factor models have been developed (e.g., Hu and Schennach (2008), Cunha, Heckman, and Schennach (2010), Gunsilius and Schennach (2019)) and come quite close to handling a fully general nonlinear measurement system $X = m(X^*, \Delta X)$.

Although completely general nonclassical measurements are notoriously difficult to handle, more specific classes of nonclassical measurements have proven empirically and methodologically useful.

One example is the so-called Berkson-type measurement error (Berkson (1950)), which almost obeys a classical-like error structure, but with the role of the measurement and the true value reversed:

Definition 2 *A measurement X (or an error ΔX) is said to be Berkson if*

$$X^* = X + \Delta X. \tag{4}$$

where, alternatively, (i) ΔX is independent from X and $E[\Delta X] = 0$ ("strongly Berkson") or (ii) $E[\Delta X|X = x] = 0$ ("weakly Berkson").

In this setting, we generally have that $X - X^*$ is not independent from X^* and that $E[X - X^*|X^*] \neq 0$, hence the nonclassical nature. This error model has traditionally been used when one has imperfect control over a given variable (e.g. medication dosage or oven temperature) and the actual resulting value of that variable deviates randomly from the imposed value. The importance of Berkson errors is beginning to be recognized in economics. For instance, it has been argued to arise 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)) or when averaged regressors and "regional" variables are used (Schennach (2013b), Blundell, Horowitz, and Parey (2018)). Rounding (Hoderlein, Siflinger, and Winter (2015)) can also generate errors that closely mimic Berkson errors, although the error $X^* - X$ is generally not completely independent from X in that case. Classical and Berkson errors can even occur simultaneously (Carroll, Delaigle, and Hall (2007), Delaigle (2007), Mallick, Hoffman, and Carroll (2002), Stram, Huberman, and Wu (2002), Hyslop and Imbens (2001)).

Another specific form of nonclassical errors was proposed by Hu and Schennach (2008).

Definition 3 *A measurement X and its true unobserved counterpart X^* satisfy a centering restriction if*

$$M\left[f_{X|X^*}(\cdot|x^*)\right] = x^* \tag{5}$$

for all x^*, where M is a known user-specified functional mapping a density on $\mathbb{R}^{d_X}$ to a point in $\mathbb{R}^{d_{X^*}}$.

Examples of such functionals include the median, the mode or the mean (in the latter case, this condition reduces to the usual weakly classical assumption). It is important to note that methods that allow for multiple measurements of the same variable with nonclassical errors typically only require the centering restriction to hold

for one of the measurements (so X in Definition 3 could be a subset of all available measurements). The ability to specify centering concepts other than the mean enables considerable freedom that has significant practical relevance. In fact, this property has been shown to hold for both the median (Bollinger (1998)) and the mode (Chen, Hong, and Tarozzi (2008)) in the well-know household income validation sample of Bound and Krueger (1991). More fundamentally, median or mode centering restrictions exhibit a robustness to other data problems not enjoyed by the mean. The mode is unaffected by truncation, while the median is unaffected by censoring, provided the truncation or censoring points do not go beyond the "center" of the distribution. One could also consider Berkson error with centering restrictions (Schennach (2013b)), i.e. $M\left[f_{X^*|X}\left(\cdot|x\right)\right] = x$.

The notion of nonclassical measurement system is so broad that it includes a number of concepts that have historically appeared under different names.

Instruments are widely used in economic applications and can be seen as special cases of indicators, often exhibiting bias or Berkson errors (Newey (2001), Schennach (2004b), Schennach (2007a), Wang and Hsiao (2011), Nadai and Lewbel (2016))

Time series (X_t) and panel data (X_{it}) often naturally provide either repeated measurements, instruments or more general indicators (Griliches and Hausman (1986), Hausman and Taylor (1981), Evdokimov (2009), Horowitz and Markatou (1996), Wilhelm (2015), Evdokimov and Zeleneev (2020a)).

The idea is that if the underlying unobservable X_t^* follows a Markov process, future ($X_{t+\ell}$) and/or past ($X_{t-\ell}$) measurements provide indirect information regarding X_t^* and can thus be considered as multiple measurements, albeit with a nonclassical error (as will be illustrated in Section 4.3). Such models have a close connection with dynamic state space models (e.g., Cunha, Heckman, and Schennach (2010), Hu and Shiu (2013), Connault (2014)), where observed variables provide indicators of the latent "state" of the system, and repeated game-theoretic models (e.g. Hu and Shum (2013)), where player's actions provide indicators of the players' valuations and/or which equilibrium they are selecting.

Mixture models (e.g. Bonhomme, Jochmans, and Robin (2016b), Hall and Zhou (2003), Compiani and Kitamura (2016), Adams (2016), Sasaki (2015a), Henry, Kitamura, and Salanié (2014), Kitamura (2003), d'Haultfoeuille and Février (2015), Jochmans, Henry, and Salanié (2017), Hu and Xiao (2018)) can be seen as special cases of repeated nonclassical measurements models where the true unobserved variable can only take on discrete values, as surveyed in more detail in Schennach (2018).

2.5 Latent variables models

Some models include unobserved (or latent) variables whose only effect is to influence the value of various observed variables. Formally, such setups can be seen as nonclassical measurement systems, although the observed variables were not *a priori* designed for specifically measuring these unobserved variables. For instance, a consumer's decision to purchase a certain type of goods does not attempt to measure their taste preferences but may very well be related to it. The idea is that, in many settings, it is possible to recover some features of the unobservables if we have access to sufficiently many observed variable related to them.

Latent variables are ubiquitous in advanced structural models. In models with unobserved heterogeneity, the latent variable X^* can represent the individual "type", or some unobserved shock that is common to multiple observations. Within a panel data setting, one could have an outcome X_{it} that is related to some observed covariates Z_{it} and to some latent "type" or common shock X_i^* and some observation-specific disturbance ΔX_{it} :

$$X_{it} = g\left(Z_{it}, X_i^*, \Delta X_{it}\right).$$

The indices would typically represent individuals (i) and time (t), but other interpretations are possible. The key to identification is that the unobserved variable keeps the same value X_i^* for different observations (X_{it}, Z_{it}) for $t = 1, \ldots, T$, hence one effectively has access to multiple measurements of a single underlying variable.

Sometimes the outcome variable X_{it} varies discontinuously with the explanatory variables (Z_{it}, X_i^*), making the identification problem more challenging, due to the resulting strong nonlinearity. A typical example comes from revealed preferences models (among many others, McFadden (2005), Aguiar and Kashaev (2020), Afriat (1973), Varian (1982)), where the outcome variable is the good being selected, while the good's characteristics and/or the individual preferences are only partially observed. Here again the identification analysis benefits from the availability of multiple outcomes where, say, unobserved preferences X_i^* are known to remain constant because the same individual is being observed, while the available goods' characteristics Z_{it} may differ. As an extreme example, Williams (2019) shows that, in the limit of a large number of measurements, even binary measurements can enable the identification of the distribution of an underlying continuous latent variable.

Latent variables can also occur due to simple data collection issues. Important examples include limited dependent variables, e.g.

$$X = \min\left\{X^*, \bar{x}\right\}$$

where $\bar{x}$ is a fixed censoring level that may be introduced for privacy reasons or due to measurement apparatus limitations. Another example is interval-valued data (Manski and Tamer (2002)), or more generally, set-valued data:

$$X = c \text{ iff } X^* \in I_c$$

where I_c for $c = 1, \ldots, C$ are disjoint possible sets of values that respondents are asked to select from. Surveys are often phrased as multiple choice questions with brackets of values rather exact numerical values, to elicit a response even if the exact value is not precisely known by the respondent. Missing data provide another example that could benefit from formal approaches aimed at handling latent variables, as an alternative to imputation procedures in cases where a proper accounting of a missing data's variability is important for consistency.

3 Basic techniques

Defining the class of problems that demand potentially nonlinear and/or nonseparable measurement systems is only part of the solution. The section is devoted to describing the methods that actually enable researchers to uncover properties of the unobservable variables from the observed variables.

3.1 Validation data

Validation data consist of an additional auxiliary sample containing data regarding the true value of the latent variables and have often been proposed as an avenue to address the presence of measurement error (Lee and Sepanski (1995), Chen, Hong, and Tamer (2005) Chen, Hong, and Tarozzi (2008), Hu and Ridder (2012)). A classic example of validation data can be found in Bound and Krueger (1991), in which mismeasured income from the Current Population Survey was matched to administrative Social Security payroll tax records to provide a validation sample for true income.

In this section, we collect general conversion formulae that relate observable quantities to the unobservable quantities of interest and that specialize to many of the standard identities that are typically used in this context. We indicate by superscripts M and V quantities pertaining to the main sample and the validation sample, respectively. In the presentation below, the main sample is assumed to provide data on two vectors Y and X, where Y represents variables not available in the validation sample, while X represents indicators for the underlying true variables X^* of interest that are

not available in the main sample. The validation sample contains data on X and X^* (which may not have the same dimension).

The results reported below are based on the following assumptions.

Assumption 1 *(i)* $f^M_{Y|XX^*}(y|x,x^*) = f^M_{Y|X^*}(y|x^*)$ *(nondifferential errors) and either (ii)* $f^M_{X|X^*}(x|x^*) = f^V_{X|X^*}(x|x^*)$ *(classical transferability) or (ii*)* $f^M_{X^*|X}(x^*|x) = f^V_{X^*|X}(x^*|x)$ *(Berkson transferability).*

Condition 1(i) requires the mismeasured variable x to provide no more information regarding the dependent variable y than the true value x^* already does. The conditions 1(ii) or (ii*) demand that the measurement error distribution be transferable between the two samples. The two cases of (ii) and (ii*) are most natural for classical and Berkson errors, respectively, which suggest the name we have given those assumptions here. However, fully nonclassical errors are allowed in either case. Both conditions (ii) and (ii*) are implied by the stronger condition that $f^M_{X^*,X}(x^*,x) = f^V_{X^*,X}(x^*,x)$. One then has the following result:

Theorem 1 *Under Assumption 1(i),(ii),*

$$f^M_{YX}(y,x) = \int f^M_{Y,X^*}(y,x^*)\, f^V_{X|X^*}(x|x^*)\, dx^* \tag{6}$$

while under Assumption 1(i),(ii),*

$$f^M_{Y|X}(y|x) = \int f^M_{Y|X^*}(y|x^*)\, f^V_{X^*|X}(x^*|x)\, dx^* \tag{7}$$

$$E^M[Y|X=x] = \int E^M[Y|X^*=x^*]\, f^V_{X^*|X}(x^*|x)\, dx^*. \tag{8}$$

When X^* and $\Delta X \equiv X - X^*$ can be assumed independent, we also have

$$f^V_{X^*|X}(x^*|x) = \frac{f^V_{\Delta X}(x-x^*)\, f^V_{X^*}(x^*)}{f^M_X(x)}, \tag{9}$$

which is useful when the validation sample only contains data on X^* (and not X) (Hu and Ridder (2012)). Then, $f^V_{\Delta X^*}$ can be inferred from $f^M_X(x)$ and $f^V_{X^*}(x^*)$ by a deconvolution argument (see Section 3.3), while $f^M_X(x)$ and $f^V_{X^*}(x^*)$ are directly available. Equation (9) is also very useful outside of a validation data context. Hsiao (1989) uses it in the context of a fully parametric model, while Li (2002) exploits repeated measurements to recover $f_{\Delta X^*}$ and f_{X^*}, as discussed further in Section 3.4.2.

Let $Y = g(X^*, \theta)$ for some given vector of moment function $g(X^*, \theta)$ that depends the true variable X^* and a parameter vector θ of interest. This substitution leads to a very useful strategy to handle measurement error in the context of GMM estimators (Chen, Hong, and Tamer (2005), Chen, Hong, and Tarozzi (2008)), after defining a "measurement error-corrected" moment condition

$$\tilde{g}(x, \theta) \equiv E^M [g(X^*, \theta) | X = x] = \int g(X^*, \theta) f^V_{X^*|X}(x^*|x) dx^*. \tag{10}$$

Indeed, by iterated expectations, one can readily verify that $E[g(X^*, \theta^*)] = 0$ if and only if $E[\tilde{g}(X, \theta^*)] = 0$, so that the validation data can be used to calculate the function $\tilde{g}(X, \theta^*)$ via (10) and the resulting feasible moment condition $E[\tilde{g}(X, \theta^*)] = 0$ can be used to determine θ with a standard GMM estimation using the main sample (although the fact that $\tilde{g}(x, \theta^*)$ is itself estimated leads to a modified asymptotic treatment, see Chen, Hong, and Tamer (2005) and Chen, Hong, and Tarozzi (2008)).

Unfortunately, validation data are, more often than not, unavailable and, even when they are, the validation sample size tends to be small (due to the difficulty or costs associated with collecting it), thus limiting the accuracy of the resulting estimators. For this reason, there is considerable interest in devising methods that do not rely on validation data.

3.2 Sensitivity analysis

Sensitivity analysis seeks to determine the effect of the measurement error of a certain type and magnitude on an assumed model with a specific functional form, in an effort to determine what types of bias one could expect measurement error to introduce. It is relatively straightforward to perform such an exercise, for instance, for polynomials and assuming normal errors (Cheng and Schneeweiss (1998), Stoker, Berndt, Ellerman, and Schennach (2005)). Sensitivity analysis does not deliver formal identification results but its simplicity and transparency make it a useful diagnostic tool. The idea of looking at the effect of measurement error on observable quantities, without attempting to recover the properties of the true covariates has also led to useful tests for the presence of measurement error (Wilhelm (2019)).

Chesher (1991) and Chesher (2017) have derived the measurement error-induced bias in conventional regressions, quantile regression models and density estimation, in the limit of small classical measurement errors. Here we report a simple multivariate

generalization of his result for errors ΔX with a small covariance matrix Σ:

$$\begin{aligned} E[Y|X=x] &= g(x) + \frac{1}{2}\operatorname{tr}\left(\left(\nabla\nabla' g(x) + 2\left(\nabla \ln f_{X^*}(x)\right)\left(\nabla g(x)\right)'\right)\Sigma\right) \\ &\quad + o\left(\|\Sigma\|\right) \end{aligned} \tag{11}$$

$$f_X(x) = f_{X^*}(x) + \frac{1}{2}\operatorname{tr}\left(\nabla\nabla' f_{X^*}(x)\,\Sigma\right) + o\left(\|\Sigma\|\right) \tag{12}$$

$$\begin{aligned} Q_{Y|X}(\tau, x) &= Q_{Y|X^*}(\tau, x) - \frac{1}{2}\frac{\operatorname{tr}\left(\left(\nabla\nabla' F_{Y|X^*} + 2\nabla \ln f_{X^*}(x)\left(\nabla F_{Y|X^*}\right)'\right)\Sigma\right)}{f_{Y|X^*}\left(Q_X(\tau, x), x\right)} \\ &\quad + o\left(\|\Sigma\|\right) \end{aligned} \tag{13}$$

where $g(x) = E[Y|X^* = x]$, ∇ denote gradients and $\nabla\nabla'$ denote Hessians while primes denote transposition and where the quantile function $Q_{Y|V}(\tau, v)$ is defined implicitly via $F_{Y|V}\left(Q_V(\tau, v)|v\right) = \tau$ for any random variable V and $\nabla F_{Y|X^*}$ is a shorthand for $\left[\nabla F_{Y|X^*}(y|x)\right]_{y=Q_{Y|X}(\tau,x)}$, with the gradient being taken with respect to the conditioning variable (and similarly for $\nabla\nabla' F_{Y|X^*}$).

Equations (11) and (12) are entirely analogous to the bias of nonparametric kernel estimators with a second order kernel (Härdle and Linton (1994)) since the effect of measurement error is to smooth all functions via a convolution with the measurement error density, which plays the role of the kernel. Chesher (1991) also notes that one can use the approximations $g(x) \approx E[Y|X = x]$ and $f_{X^*}(x) \approx f_X(x)$, etc. in the bias expressions without affecting the order of the $o\left(\|\Sigma\|\right)$ remainders. This result is useful, because it enables the researcher, based on the observed data alone, to easily get an idea of what would be the direction and magnitude of the bias introduced by a measurement error of a given (small) magnitude (see Battistin and Chesher (2014) for an application of this approach to treatment effects and Chesher and Schluter (2002) for an application to the assessment of the impact of measurement error on welfare measures). 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 ($\nabla\nabla' g(x)$) and nonuniformities in the density of the regressor ($\nabla\left(\ln f_{X^*}(x)\right)$), while densities are only affected by curvature ($\nabla\nabla' f_{X^*}(x)$). This small-error expansion approach can be applied to GMM settings as well and, in this context, Evdokimov and Zeleneev (2020b) shows how Σ could be identified if a sufficient number of moment conditions are available.

This approach can be adapted to the Berkson case (Schennach (2016)):

$$E[Y|X=x] = g(x) - \frac{1}{2}\operatorname{tr}\left(\nabla\nabla' g(x)\,\Sigma\right) + o\left(\|\Sigma\|\right) \tag{14}$$

$$f_X(x) = f_{X^*}(x) - \frac{1}{2}\operatorname{tr}\left(\nabla\nabla' f_{X^*}(x)\,\Sigma\right) + o\left(\|\Sigma\|\right) \tag{15}$$

$$Q_{Y|X}(\tau,x) = Q_{Y|X^*}(\tau,x) + \frac{1}{2}\frac{\operatorname{tr}\left(\left(\nabla\nabla' F_{Y|X^*}\right)\Sigma\right)}{f_{Y|X^*}\left(Q_X(\tau,x),x\right)} + o\left(\|\Sigma\|\right). \tag{16}$$

While the bias contains terms similar to Equations (11)-(13), with the expected opposite sign, Equation (14) is free of terms that involve first derivatives. This reflects the fact that a conventional least squares regression of Y on a Berkson-type error contaminated regressor X is consistent in the special cases of linear or quadratic specifications (Huwang and Huang (2000), Hausman, Newey, Ichimura, and Powell (1991)), with only a biased intercept coefficient in the quadratic case.

3.3 Deconvolution

Deconvolution is a widely studied method that is unfortunately only helpful when the measurement error ΔX is strongly classical and its distribution is known. The density of the sum of two independent random variables (or random vectors) X^* and ΔX is given by the familiar convolution of their respective densities:

$$f_X(x) = \int f_{\Delta X}(x - x^*)\, f_{X^*}(x^*)\, dx^*. \tag{17}$$

This result takes an even simpler form of a conventional product when expressed in terms of Fourier transforms:

Definition 4 *For any random variable (or random vector) V, we define the* characteristic function *as $\phi_V(\xi) \equiv E\left[e^{\mathbf{i}\xi\cdot V}\right]$, where $\mathbf{i} \equiv \sqrt{-1}$.*

Theorem 2 *(Convolution theorem) For two independent random variables X^* and ΔX,*

$$\phi_{X^*+\Delta X}(\xi) = \phi_{X^*}(\xi)\,\phi_{\Delta X}(\xi). \tag{18}$$

It is useful to observe that independence is a sufficient but not a necessary condition for Equation (18) to hold. The necessary and sufficient condition is "subindependence", as defined below (Hamedani and Volkmer (2009), Ebrahimi, Hamedani, Soofi, and Volkmer (2010), Hamedani (2013), Schennach (2019)):

Definition 5 *X and Y are independent iff $\phi_{X,Y}(\xi,\gamma) = \phi_X(\xi)\,\phi_Y(\gamma)$ for all ξ,γ while they are* subindependent *iff $\phi_{X,Y}(\xi,\xi) = \phi_X(\xi)\,\phi_Y(\xi)$ for all ξ.*

Subindependence is arguably as weak as a conditional mean assumption (Schennach (2019)), because both conditions can be phrased, in Fourier representation, as a constraint on a one-dimensional subspace of frequency space, rather than a two-dimensional subspace in the case of independence. This observation is useful in econometric modeling, because it emphasizes the fact that many measurement error techniques that rely on independence are, in fact, much more robust to violations of this assumption than previously thought. The usefulness of the notion of subindependence is beginning to be more broadly recognized in econometrics (Adusumilli, Otsu, and Whang (2017), Chen, Linton, and Yi (2017)).

The main advantage of the Fourier representation is to facilitate the reverse operation of deconvolution. If the distribution of the measurement error is known, then $\phi_{\Delta X}(\xi)$ is known, and we can thus express the density of the true unobserved variables in terms of known ($\phi_{\Delta X}(\xi)$) or observed ($\phi_X(\xi)$) quantities:

$$f_{X^*}(x^*) = (2\pi)^{-d_X} \int \frac{\phi_X(\xi)}{\phi_{\Delta X}(\xi)} e^{-\mathbf{i}\xi\cdot x^*} d\xi \tag{19}$$

where the integral is taken to be over $\mathbb{R}^{d_X}$, under the standard assumption that $\phi_{\Delta X}(\xi) \neq 0$ (it is sufficient that this holds on a dense subset of $\mathbb{R}^{d_X}$, since $\phi_{X^*}(\xi)$ is continuous (Carrasco and Florens (2011))).

This identification result can be naturally turned into an estimator by taking advantage of the fact that kernel smoothing is also a type of convolution, leading to the kernel deconvolution estimator

$$\hat{f}_{X^*}(x^*) \equiv (2\pi)^{-d_X} \int \left(\frac{1}{n} \sum_{j=1}^{n} e^{\mathbf{i}\xi\cdot X_j} \right) \frac{\phi_K(h\xi)}{\phi_{\Delta X}(\xi)} e^{-\mathbf{i}\xi\cdot x^*} d\xi \tag{20}$$

where $\phi_K(\xi) = \int K(x)\, e^{i\xi x} dx$ is the Fourier transform of the kernel K while h is the bandwidth. This estimator 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). Alternative methods, not based on Fourier transforms, have also been proposed (e.g., Carrasco and Florens (2011), Mallows (2007), Arellano and Bonhomme (2020)).

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

$$\hat{E}\left[Y|X^*=x^*\right]=\frac{\int\frac{\dot{\phi}_X^Y(\xi)}{\phi_{\Delta X}(\xi)}\phi_K\left(h\xi\right)e^{-\mathbf{i}\xi\cdot x^*}d\xi}{\int\frac{\dot{\phi}_X^1(\xi)}{\phi_{\Delta X}(\xi)}\phi_K\left(h\xi\right)e^{-\mathbf{i}\xi\cdot x^*}d\xi} \tag{21}$$

where we have introduced the quantity $\dot{\phi}_X^Y(\xi)\equiv\frac{1}{n}\sum_{j=1}^n Y_j e^{\mathbf{i}\xi\cdot X_j}$, which can be viewed as an estimator of $E\left[Ye^{\mathbf{i}\xi\cdot X}\right]$.

A notable difficulty in applying kernel deconvolution is that the division by $\phi_{\Delta X}(\xi)$ in Equations (20) and (21) can cause the integrand to become very noisy in the limit of large frequency ξ. Keeping this unwanted phenomenon under control typically demands (i) a kernel whose Fourier transform $\phi_K(\xi)$ has compact support and (ii) a slower decrease of the bandwidth h with sample size n, relative to standard nonparametric estimators. The latter often results in slow convergence (in some cases, as slow as $(\ln n)^{-\alpha}$ for some $\alpha>0$).

Although kernel deconvolution methods for classical errors are often criticized for their slow convergence, an arguably more severe limitation is their reliance on the knowledge of the measurement error distribution. Fortunately, numerous alternative methods, to be discussed in the subsequent sections, have been proposed that do not share this limitation.

One can also similarly handle the converse case of Berkson (rather than classical) errors using Fourier techniques (Delaigle, Hall, and Qiu (2006)). In fact, this case is arguably much simpler and convenient because (i) Fourier transforms are not actually necessary, as Equation (17), with the role of X and ΔX reversed, suffices and (ii) the ill-posed operation of dividing by a potentially small $\phi_{\Delta X}(\xi)$ is avoided. The end result is that, if the measurement error distribution is known, one can obtain rapidly converging (even root n consistent) estimators.

3.4 Repeated measurements

Repeated measurements arguably provide the most practical way to address measurement error problems. They consist of multiple imperfect measurements of the same unobserved underlying true variable (Hausman, Newey, Ichimura, and Powell (1991), Li and Vuong (1998), Li (2002), Schennach (2004a), Schennach (2004c), Delaigle, Hall, and Meister (2008)). This class of approaches proves very useful when there are no biases in at least one of the measurements, because then, as we will see, there exist formal methods to fully recover the distribution or various features (e.g. moments) of

the underlying true variable from the observed distributions. A well-known application of this approach is the estimation of so-called Engel curves that relate household expenditure share on various types of good to household income, all of which are potentially mismeasured (Hausman, Newey, and Powell (1995), Schennach (2004a)). As consumption or income data is often collected over time, observations in different time periods can serve as repeated measurements.

Although we shall focus on the minimum number of measurement needed for identification in this review, the availability of more measurements can be straightforwardly exploited to improve statistical accuracy by merely averaging over the different estimates obtained while considering different subsets of the measurements.

3.4.1 Correlation analysis

Correlations are mostly useful in linear models. As a simple example, consider three scalar measurements $X = X^* + \Delta X$, $Y = \alpha X^* + \Delta Y$ and $Z = \beta X^* + \Delta Z$ with $X^*, \Delta X, \Delta Y, \Delta Z$ mutually uncorrelated and with zero mean. One can then straightforwardly identify the model's key parameters through:

$$\begin{array}{cc} \alpha = E[YZ]/E[XZ] & \beta = E[YZ]/E[YX] \\ \mathrm{Var}[X^*] = E[YX]/\alpha = E[ZX]/\beta & \mathrm{Var}[\Delta X] = \mathrm{Var}[X] - \mathrm{Var}[X^*] \\ \mathrm{Var}[\Delta Y] = \mathrm{Var}[Y] - \alpha^2 \mathrm{Var}[X^*] & \mathrm{Var}[\Delta Z] = \mathrm{Var}[Z] - \beta^2 \mathrm{Var}[X^*] \end{array} \tag{22}$$

Generalizing of this basic idea leads to a *factor model* structure between the observed vector of measurements X and the unobserved vector of *factors* X^*:

$$X = \Lambda X^* + \Delta X \tag{23}$$

where $E[\Delta X] = 0$, $E[X^*] = 0$, $\mathrm{Covar}[X^*, \Delta X] = 0$ and $\mathrm{Var}[\Delta X]$ is diagonal while Λ is a $d_X \times d_{X^*}$ matrix of *factor loadings* (here $d_X \neq d_{X^*}$ in general). The above assumes, without loss of generality, that X has been demeaned.

Such model can be shown to be identified (up to some normalizations) under simple conditions (Anderson and Rubin (1956)):

Theorem 3 *In model (23), if the matrix Λ is such that there remains two disjoint matrices of rank d_{X^*} after any one row of Λ is removed, then a given $\mathrm{Var}[X]$ uniquely determines $\mathrm{Var}[\Delta X]$, while all observationally equivalent Λ have the form $\tilde{\Lambda}T$ for some fixed $\tilde{\Lambda}$ and an arbitrary $d_{X^*} \times d_{X^*}$ matrix T satisfying $T'T = I$.*

Normalizations are necessary in this model, because it is possible to substitute $\Lambda = \tilde{\Lambda}T^{-1}$ and $X^* = T\tilde{X}^*$ into (23), without affecting the observable quantities, while

obtaining different factors $\tilde{X}^*$ and different factor loadings $\tilde{\Lambda}$. A popular normalization is to assume that $\text{Var}[X^*] = I$.

Alternatively one may assume that d_{X^*} of the measurements in X are each dedicated to only one latent factor (so that Λ contains, on a diagonal block, a submatrix that is a $d_{X^*} \times d_{X^*}$ identity matrix). In this case, the elements of X^* can be generally correlated. It is interesting to note that the selection of a specific normalization becomes unnecessary when the latent factors X^* are merely used as control variables (for instance, to match comparable individuals in treatment effects models).

The rank condition of Theorem 3 is sufficient but not necessary in general. While identification of Λ from the covariances and normalizations alone clearly requires $d_X \geq 2d_{X^*} + 1$ to satisfy the rank condition of Theorem 3, that number can be reduced using higher order moments (Bonhomme and Robin (2009)), or, equivalently, higher order derivatives of the log characteristic function (Ben-Moshe (2018)).

An analysis of the correlation structure is also very useful in time series or panel data settings (Griliches and Hausman (1986), Gospodinov, Komunjer, and Ng (2017)). If one posits distinct models for the serial correlation of the signal and the noise, it is often possible to recover the model's primitives (e.g. autocorrelation or moving average coefficients) from the observed correlations over time by solving a system of nonlinear equations.

3.4.2 Kotlarski-type identities

Consider two repeated measurements X, Z that are related to the true underlying variable of interest X^* via

$$X = X^* + \Delta X \tag{24}$$
$$Z = X^* + \Delta Z \tag{25}$$

where the appropriate assumptions regarding the measurement errors ΔX and ΔZ will be specified for each case below. 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)):

Lemma 1 *In the system (24)-(25), 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 $E\left[e^{\mathbf{i}\zeta Z}\right] \neq 0$ for all real ζ,*

$$\phi_{X^*}(\xi) = \exp\left(\int_0^\xi \frac{E\left[\mathbf{i}Xe^{\mathbf{i}\zeta Z}\right]}{E\left[e^{\mathbf{i}\zeta Z}\right]} d\zeta\right) \tag{26}$$

and $f_{X^*}(x^*) = (2\pi)^{-1} \int \phi_{X^*}(\xi) e^{-\mathbf{i}\xi x} d\xi$.

Kotlarski's Lemma has been modified and generalized in various ways since then. 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 settings (assuming mutual independence of X^*,ΔZ_1,...,$\Delta Z_{d_{X^*}}$):

$$\begin{aligned} \phi_{X_k^*}(\xi_k) &= \exp\left(\int_0^{\xi_k} \frac{E\left[\mathbf{i}X_k e^{\mathbf{i}\zeta Z_k}\right]}{E\left[e^{\mathbf{i}\zeta Z_k}\right]} d\zeta\right) \\ \phi_{X^*}(\xi) &= \frac{\phi_Z(\xi)}{\prod_{k=1}^{d_{X^*}} \left(\phi_{Z_k}(\xi_k) / \phi_{X_k^*}(\xi_k)\right)} \end{aligned}$$

Li (2002) also uses this result to correct for measurement error in a nonlinear regression model by combining Equations (26) and (9).

Schennach (2004a) obtains an identity that generalizes both Kotlarski's result and the identification result for polynomial errors-in-variables regression model of Hausman, Newey, Ichimura, and Powell (1991).

Theorem 4 *Let* X^*, X, Y, Z *be random vectors, with* $d_{X^*} = d_X = d_Z$. *If (i)* $E[\Delta X | X^*, \Delta Z] = 0$, *(ii)* $E[Y|X^*, \Delta Z] = E[Y|X^*]$,[1] *(iii)* ΔZ *is independent from* X^*, *then, for any function* $u(x^*)$ *with Fourier transform* $\mu(\xi)$,

$$E[Yu(X^*)] = (2\pi)^{-d_X} \int_{\mathbb{R}^{d_X}} \mu(-\xi) \frac{E\left[Y e^{\mathbf{i}\xi \cdot Z}\right]}{E\left[e^{\mathbf{i}\xi \cdot Z}\right]} \phi_{X^*}(\xi) \, d\xi \tag{27}$$

where the integral is over the whole $\mathbb{R}^{d_X}$ *space and*

$$\phi_{X^*}(\xi) = \exp\left(\int_0^{\xi} \frac{E\left[\mathbf{i}X e^{\mathbf{i}\zeta \cdot Z}\right]}{E\left[e^{\mathbf{i}\zeta \cdot Z}\right]} \cdot d\zeta\right) \tag{28}$$

where the integral is the path integral of a vector-valued field along a 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).*

In this framework, Y collect variables that enter the moments linearly, which can exhibit zero mean errors conditional on X^* (such as the dependent variable in a least-square regression). If Y is not needed, it can be simply set to $Y = 1$. Perfectly

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

measured variables that enter the moment nonlinearly can be included in X^* and the corresponding elements of X and Z can then just be set to be equal.

This result is phrased in a form that directly identifies moments involving (a possibly multivariate) X^* (and perhaps other variables Y), because a large class of models can be identified from the knowledge of a set of moments. This is true by construction for GMM-type models, but also for likelihood models. As a special case, for nonlinear regression models:

$$Y = g\left(X^*, \theta\right) + \Delta Y \tag{29}$$

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

The Fourier transform $\mu\left(\xi\right)$ may be a generalized function (Lighthill (1962), Schwartz (1966), Temple (1963), Gel'fand and Shilov (1964)). For instance, if $u\left(x^*\right)$ is a polynomial, the Fourier transform $\mu\left(\xi\right)$ consists of delta function derivatives of various orders that effectively extract various derivatives of the quantity $\left(E\left[Ye^{\mathbf{i}\xi\cdot Z}\right] / E\left[e^{\mathbf{i}\xi\cdot Z}\right]\right)\phi_{X^*}\left(\xi\right)$ in (27). In this fashion, one can recover the polynomial result of Hausman, Newey, Ichimura, and Powell (1991) for repeated measurements under the same conditional mean and independence assumptions. This observation is especially useful in multivariate settings, where a polynomial result has not yet been explicitly derived (and would be extremely unwieldy), but a symbolic algebra engine can automatically compute the appropriate mixed derivative of $\left(E\left[Ye^{\mathbf{i}\xi\cdot Z}\right] / E\left[e^{\mathbf{i}\xi\cdot Z}\right]\right)\phi_{X^*}\left(\xi\right)$ and Equation (28).

Equations (27) and (28) also suggest a very natural estimator in which all quantities of the form $E\left[Ve^{\mathbf{i}\zeta\cdot Z}\right]$ for $V = 1, Y$ are replaced by sample averages $n^{-1}\sum_{j=1}^{n} V_j e^{\mathbf{i}\zeta\cdot Z_j}$. Schennach (2004a) shows that, after a simple automatic bounding device (ensuring, for instance, that an estimator obtained via (28) is bounded), this approach yields a root n consistent and asymptotically normal estimator that does not require any user-specified bandwidth parameter, provided $u\left(x^*\right)$ is sufficiently smooth. This smoothness condition ensures that $\mu\left(\xi\right)$ decays sufficiently rapidly as $|\xi| \to \infty$ to downweigh the noise in the tail in the estimated Fourier transform, yielding a finite overall noise that decays to zero at the rate $n^{-1/2}$ as $n \to \infty$.

One can also recover Kotlarski's identity from (27) and (28) under fewer independence assumptions by setting $Y = 1$ and $\mu\left(\xi\right) = e^{\mathbf{i}\xi\cdot x_0^*}$ (which corresponds to setting $u\left(x^*\right)$ to be a delta function, or a point mass, at x_0^*). In a similar vein, only setting $\mu\left(\xi\right) = e^{\mathbf{i}\xi\cdot x_0^*}$ (but keeping Y) yields the identification of $E\left[Y|X^* = x_0^*\right] f_{X^*}\left(x_0^*\right)$ which opens the way to nonparametric identification of conditional expectations (after division by $f_{X^*}\left(x_0^*\right)$, which is also identified):

$$E[Y|X^*=x] = \frac{\int e^{-\mathbf{i}\xi\cdot x_0^*} \frac{E\left[Ye^{\mathbf{i}\xi\cdot Z}\right]}{E\left[e^{\mathbf{i}\xi\cdot Z}\right]} \phi_{X^*}(\xi)\, d\xi}{\int e^{-\mathbf{i}\xi\cdot x_0^*} \phi_{X^*}(\xi)\, d\xi} \tag{30}$$

with $\phi_{X^*}(\xi)$ given by (28). Schennach (2004c) proposes a nonparametric estimator based on this identity while Schennach, White, and Chalak (2012) also considers the estimation of derivatives as well as functionals constructed from such quantities. Similar constructs have found applications for varying coefficient models (Dong, Otsu, and Taylor (2020)).

Kotlarski's identity has been generalized in other ways. For instance, in order to relax some of the nonvanishing Fourier transform assumptions, Evdokimov (2009) 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. This only requires a nonvanishing assumption on $\phi_{\Delta Z}(\xi)$ but not $\phi_{X^*}(\xi)$. This can be even further relaxed by evaluating an appropriate limiting process, provided higher order derivatives of $\phi_{X^*}(\xi)$ do not vanish where $\phi_{X^*}(\xi)$ itself does (see Schennach (2000) and Evdokimov and White (2012)). One should note, however, that distributions whose characteristic functions vanish are relatively rare: Among commonly used distributions, the uniform and the symmetric triangular distributions are the only examples.

The structure of the second measurement can also be extended to $Z = a+bX^*+\Delta Z$ with $E[\Delta Z] = 0$ if another variable Y, related to X^* but independent from the errors, is available (Carroll, Ruppert, Crainiceanu, Tosteson, and Karagas (2004)). In this case, the slope coefficient can be identified from $b = \mathrm{Cov}(Y,Z) / \mathrm{Cov}(Y,X)$ while the intercept is given by $a = E[Z] - bE[X]$. Hu and Sasaki (2015) were able to further extend Kotlarski's approach to a measurement system that includes a measurement that is polynomially related to its true unobserved counterpart, thus allowing for the possibility of a nonclassical measurement.

Kotlarski-type identities can be applied to a class of factor models (in which the indicators may not be obviously matched with corresponding latent factors):

$$X = \Lambda X^* + \Delta X \text{ with } X^* \perp\!\!\!\perp \Delta X \tag{31}$$

where $\perp\!\!\!\perp$ denotes independence and where we may assume without loss of generality that all variables have zero mean. Let us assume that the factor loading matrix Λ is known (as it can be identified, e.g., from Theorem 3). One can then construct two vectors of repeated measurements suitable for use in Equation (28) in order to estimate the joint distribution of the factors X^*. Specifically, under the same conditions as Theorem 3 and under a suitable normalization, one can decompose the known 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 (31). Then, the repeated measurement vectors X and Z in Equation (28) can be taken to be $X \equiv \Lambda_A^{-1} X_A$ and $Z \equiv \Lambda_B^{-1} X_B$.

Bonhomme and Robin (2010) instead consider a factor model of the form

$$X = \Lambda X^* \text{ with } X_k^* \perp\!\!\!\perp X_\ell^* \text{ for } k \neq \ell \text{ and } E\left[X_k^*\right] = 0 \tag{32}$$

In this framework, one combines all unobservables in one vector without labelling some as noise and others as signal. The matrix of coefficients Λ is again considered known. They show that the marginal distribution of each X_k^* is identified using a Kotlarski-type identity that involves second derivatives of Fourier transforms.

Theorem 5 *In Model (32),*

$$\ln \phi_{X_k^*}(\chi) = \int_0^{\chi} \int_0^{\zeta} \left(Q^-\right)_{k\cdot} \operatorname{vech}\left(\nabla\nabla' \ln \phi_X\left(\frac{\xi\theta}{\theta' \Lambda_{\cdot k}}\right)\right) d\xi d\zeta$$

where $\theta \in \mathbb{R}^{d_X} \setminus \{0\}$ is a user-specified direction of integration and where $Q^- = \left(Q^T Q\right)^{-1} Q^T$ and $Q = \left(\operatorname{vech}\left(\Lambda_{\cdot 1}\left(\Lambda_{\cdot 1}\right)'\right), \ldots, \operatorname{vech}\left(\Lambda_{\cdot d_{X^*}}\left(\Lambda_{\cdot d_{X^*}}\right)'\right)\right)$ provided Q has full column rank d_{X^*} and $\phi_{X_k^*}(\chi)$ does not vanish

The requirement of independent factors can be relaxed to some extent. Ben-Moshe (2018) provides a range of identification theorems allowing for various levels of dependence between the factors, from general dependence to mean independence. His results, many of which are necessary and sufficient, are phrased in terms of first and/or second derivatives of the characteristic functions.

Kotlarski-type identities are being increasingly used in panel data settings (Evdokimov (2009), Arellano and Bonhomme (2011)), based on the idea that the "fixed" or "permanent" effects remain constant over time while transitory components do not, so that observations over time can provide a form of repeated measurements.

Kotlarski-type identities have also been modified in rather different directions, to handle nonlinear measurement error problems in a variable X^* with an instrument vector (denoted W), rather than a conventional repeated measurement:

$$Y = g\left(X^*, \Delta Y\right) \tag{33}$$

$$X = X^* + \Delta X \tag{34}$$

$$X^* = h\left(W\right) + \Delta Z. \tag{35}$$

For convenience, one can work with the constructed instrument Z, defined as $Z \equiv h(W)$, since the function $h(\cdot)$ is identified from observables via $h(w) = E[X|W = w]$.

The proof of identification of this model in the separable case $Y = g(X^*) + \Delta Y$ with scalar X^* has a long history, starting from the polynomial treatment of Hausman, Newey, Ichimura, and Powell (1991), followed by a conjecture, based on functional equations, that identification holds more generally (Newey (2001)). Some counterexamples based on exponential specifications were described in Schennach (2004b), while Wang and Hsiao (2011) showed identification in specific parametric models satisfying integrability conditions. Schennach (2007a) then showed identification in general settings by using Fourier transforms of the functional equations of Newey (2001). The resulting expression for $\phi(\zeta)$ is similar to Kotlarski's result except for two key differences: (i) all 3 variables Y, X and Z (rather than 2 of them) play a role in identification and (ii) the Fourier transforms involve conditional expectations rather than probability densities. As conditional expectations are not necessarily integrable, their Fourier transform need to be interpreted as generalized functions (Lighthill (1962), Schwartz (1966), Temple (1963)). Nadai and Lewbel (2016) have extended this approach by (i) allowing for some forms of correlations between the errors in Y and X, (ii) considering multiplicative errors and (iii) providing identification results for polynomial moments of Y.

Schennach (2008) extends Schennach (2007a) to quantile regressions, thus allowing for a nonseparable model of the form $Y = g(X^*, \Delta Y)$ and enabling the recovery of the complete joint distribution of Y and X^*. Remarkably, this result can be phrased entirely in terms of ordinary (rather than generalized) functions, thus suggesting a natural plug-in estimator. An extension of this result to multivariate X^* (and thus multivariate X and Z), developed by Wilhelm (2015), is reported below.

Theorem 6 *In Model (33)-(35), if (i) Z is supported on all of $\mathbb{R}^{d_{X^*}}$, (ii) ΔX^* admits a bounded density and $E[\|\Delta X^*\|] < \infty$, (iii) $Y \perp\!\!\!\perp W \mid X^*$, $\Delta X^* \perp\!\!\!\perp W$ and $E[\Delta X|W = w, Y = y] = 0$ and (iv) there exist $B < \infty$ and $\beta > 0$ such that*

$$\sup_q \left| \frac{\partial^{d_{X^*}} P[Y \leq q|X^* = x^*]}{\partial x_1^* \ldots \partial x_{d_{X^*}}^*} \right| \leq B \prod_{j=1}^{d_{X^*}} \left(1 + \left|x_j^*\right|\right)^{-3-\beta} \tag{36}$$

for all $x^* \in \mathbb{R}^{d_{X^*}}$, then, the cdf of Y given X^* is given by

$$P[Y \leq q|X^* = x^*] = c(q) + (2\pi)^{-d_{X^*}} \int \frac{\sigma(\zeta, q)}{\phi(\zeta)} e^{-\mathrm{i}\zeta \cdot x^*} d\zeta \tag{37}$$

where, for a given $q_0 \in \mathbb{R}$*,*

$$\sigma(\zeta, q) = \int \left(E\left[\mathbf{1}\left\{Y \leq q_0\right\} | Z = z\right] - c(q)\right) e^{\mathbf{i}\zeta \cdot z} dz \tag{38}$$

$$c(q) = \lim_{r_1 \longrightarrow \infty} \lim_{r_2 \longrightarrow \infty} \frac{\int_{r_1 \leq |z| \leq r_2} E\left[\mathbf{1}\left\{Y \leq q_0\right\} | Z = z\right] dz}{\int_{r_1 \leq |z| \leq r_2} dz} \tag{39}$$

$$\phi(\zeta) = \exp\left(\int_0^{\zeta} \frac{\mathbf{i}\sigma_1(\xi)}{\sigma_0(\xi)} \cdot d\xi\right) \tag{40}$$

$$\sigma_k(\zeta) = \int s_k(z) e^{\mathbf{i}\zeta z} dz \tag{41}$$

$$s_1(z) = \frac{\partial}{\partial z} E\left[(Z - X)\mathbf{1}\left\{Y \leq q_0\right\} | Z = z\right] \tag{42}$$

$$s_0(z) = \frac{\partial}{\partial z} E\left[\mathbf{1}\left\{Y \leq q_0\right\} | Z = z\right], \tag{43}$$

in which the denominators $\phi(\zeta)$ *and* $\sigma_0(\xi)$ *are assumed nonvanishing.*

It should be noted that it is of course possible for a dataset to exhibit both measurement error and endogeneity at the same time. In general, in a nonlinear model, a single instrument is unfortunately not sufficient to handle both problems simultaneously and dedicated methods that address each problem individually have to be used (Otsu and Taylor (2016), Song, Schennach, and White (2015), Schennach, White, and Chalak (2012)).

3.4.3 Differencing

A special case that turns out to be especially simple to handle occurs when X^*,ΔX and ΔZ are mutually independent in the repeated classical measurements setup of Equations (24) and (25) where the distribution of the measurement errors ΔX and ΔZ are identical and symmetric about zero. Then, one can exploit the fact that $X - Z$ only depends on the measurement errors (see Horowitz and Markatou (1996), Li and Vuong (1998), Delaigle, Hall, and Meister (2008)) to write $\phi_{Z-X}(\xi) = \phi_{\Delta Z - \Delta X}(\xi) = \phi_{\Delta X}(\xi)\phi_{\Delta X}(-\xi) = |\phi_{\Delta X}(\xi)|^2$ so that $\phi_{X^*}(\xi) = \phi_X(\xi) / \sqrt{|\phi_{Z-X}(\xi)|}$ (selecting the positive root, assuming $\phi_{\Delta X}(\xi)$ is strictly positive).

Delaigle and Meister (2007) provide an alternative expression that does not require the distributions of ΔX and ΔZ to be identical. Here we remark that this expression remains valid under even weaker conditions, leading to the following result (shown in Schennach (2018)):

Theorem 7 *Let Z, X and X^* each take value in $\mathbb{R}^{d_{X^*}}$. If X^* is (sub)independent from $(\Delta X + \Delta Z)/2$ and the distribution of ΔX conditional on ΔZ is symmetric about 0 then*

$$\phi_{X^*}(\xi) = \frac{\phi_{(Z+X)/2}(\xi)}{\phi_{(Z-X)/2}(\xi)}, \tag{44}$$

provided $\phi_{(Z-X)/2}(\xi) \neq 0$ for $\xi \in \mathbb{R}^{d_{X^*}}$ (except perhaps at isolated points).

3.4.4 Higher-order moments

It can happen that a measurement system is such that the dimension of X is too small relative to that of X^*, so that not enough repeated measurements are available. In such cases, one typically needs to make stronger independence assumptions regarding the observables and attempt to regain identification by exploiting every possible implications of these additional assumptions.

In a linear regression of Y on X^* :

$$Y = X^*\theta + \Delta Y \tag{45}$$

$$X = X^* + \Delta X \tag{46}$$

with X being a classical error-contaminated measurement of X^*, 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, some of which are zero by assumption. The resulting expressions can be solved for the slope parameter θ in terms of observable moments. 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), Erickson and Whited (2002), Erickson and Whited (2012), Erickson, Jiang, and Whited (2014), Dagenais and Dagenais (1997), Lewbel (2012), Lewbel (1997), Bonhomme and Robin (2009), and the many references therein) have exploited analogous techniques to identify and estimate linear errors-in-variables models without any auxiliary information.

One simple result of this kind is that if $E\left[\left(X^* - E[X^*]\right)^3\right] \neq 0$ then

$$\theta = \frac{\mathrm{Cov}[Z, Y]}{\mathrm{Cov}[Z, X]} \tag{47}$$

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

$E\left[(X^* - E[X^*])^3\right] \neq 0$ by using fourth order mixed moments, but then the restriction $E\left[(X^* - E[X^*])^4\right] \neq 3E\left[(X^* - E[X^*])^2\right]$ is required. Violations or near-violations of such nonzero denominator conditions can be readily detected in practice by inspecting the magnitude of the denominator of Equations such as (47) or, equivalently, by checking if the estimated standard errors of a corresponding estimator of θ are reasonably small. 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 Fourier transforms.

Going beyond a simple univariate regression, Bonhomme and Robin (2009) generalizes the classic treatment of Geary (1942) to multivariate factor models (introduced in Equation (23) above), enabling identification of the factor loadings with fewer measurements than from covariance information alone (Anderson and Rubin (1956)).

While using specific moments is simple and convenient, it carries the risk of projecting out some useful information. In response to this, researchers have begun using the full distribution of the observable variables instead. Chen, Hu, and Lewbel (2009) use this idea to identify a regression model with discrete mismeasured regressors. In that case, thanks to the discrete nature of the problem, the identification can be reduced to a finite, albeit complex, system of equations. In the case of continuous variables, the classic result of Reiersol (1950), covering a univariate linear regression with classical errors in both variables, has been generalized to a nonparametric univariate regression model by Schennach and Hu (2013). More recently, the case where the regressor exhibits centered nonclassical errors (while the dependent variable error remains classical) has also been shown to be identified (Hu, Schennach, and Shiu (2020)), albeit under some stronger assumptions (such as boundedness and monotonicity of the regression function).

An influential empirical example of the use of the higher-order moment approach can be found in Erickson and Whited (2000), who re-examine the apparent failure of the so-called q theory (Brainard and Tobin (1968), Tobin (1969)), which relates a firm's investment behavior to q, defined as the ratio of the market values of the firm's capital stock to its replacement value. The authors find that q theory has, in fact, good predictive power for investments, once measurement error in q is accounted for. In the absence of credible instruments, higher-order moments techniques were crucial to establish the result.

3.5 Nonclassical repeated measurements

3.5.1 Operator inversion

In the case of nonclassical errors, the density of the sum X of two random variables X^* and ΔX that are not necessarily independent is no longer given by a simple convolution, but is instead given by

$$f_X = L_{X|X^*} f_{X^*}, \tag{48}$$

where we have introduced the operator

$$\left[L_{X|X^*} f\right](x) \equiv \int f_{X|X^*}(x|x^*) f(x^*) \, dx^* \tag{49}$$

acting on a generic function f. This operator point of view makes a direct connection to the problem of inverting an integral equation (Carrasco, Florens, and Renault (2005)), which is well-studied in the nonparametric instrumental variable literature (Newey and Powell (2003), Darolles, Florens, and Renault (2011), Hall and Horowitz (2005), Adusumilli and Otsu (2015)). Note that Equation (49) is the continuous analogue of matrix multiplication, with $f_{X|X^*}$ playing the role of a matrix and f of a vector.

We assume (for this section only) that the density $f_{X|X^*}$ is known and consider the problem of recovering f_{X^*} from f_X. Whether Equation (48) can be inverted as $f_{X^*} = L^{-1}_{X|X^*} f_X$ depends on whether the operator $L_{X|X^*}$ is *injective* (Carrasco, Florens, and Renault (2005)), a condition often stated as *invertibility* (as in Hall and Horowitz (2005)) or by stating that the underlying density $f_{X|X^*}$ is *complete* or *forms a complete family* (as in Newey and Powell (2003)). Sufficient conditions for injectivity can be found in Newey and Powell (2003), Mattner (1993), d'Haultfoeuille (20011), Hu and Ridder (2010), Andrews (2017) and Hu, Schennach, and Shiu (2017), but the literature is far from a simple and exhaustive characterization of injective linear operators in terms of primitive conditions.

Even when the inverse of an operator exists, it may still lead to an ill-posed problem (i.e. the output f_{X^*} is not a continuous function of the input f_X). While this does not preclude identification, it does have implications for estimation purposes (i.e. slow convergence). One can reasonably assume that a matrix has no zero eigenvalue, but virtually all reasonable invertible integral operators have eigenvalues that are not bounded away from zero, which is the source of the problem. In practice, this means that the inverse operator must be "regularized" to obtain consistent estimators (as covered extensively in Carrasco, Florens, and Renault (2005)).

3.5.2 Operator diagonalization

Identifying nonclassical measurement error models involves more than merely inverting a known operator $L_{X|X^*}$, as it also requires the identification of this operator solely from the information provided by the observable variables. Hu and Schennach (2008) establish the identification of general nonlinear factor model with continuous variables contaminated by nonclassical measurement errors. For convenience, the vector of observable indicators is partitioned as (X, Y, Z) and the measurement system has the form

$$V = m_V\left(X^*, \Delta V\right) \text{ for } V = X, Y, Z. \tag{50}$$

One can think of Y as the dependent variable and X^* as the regressor for which two indicators X and Z are available, but this assignment is not the only possible choice.

Theorem 8 *Assume that (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*[2] *(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 $M\left[f_{X|X^*}\left(\cdot|x^*\right)\right] = x^*$. Then, for a given true observed density $f_{YX|Z}$, the equation*

$$f_{YX|Z}\left(y, x|z\right) = \int f_{Y|X^*}\left(y|x^*\right) f_{X|X^*}\left(x|x^*\right) f_{X^*|Z}\left(x^*|z\right) dx^* \tag{51}$$

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

$$f_{YXZ}\left(y, x, z\right) = \int f_{X|X^*}\left(x|x^*\right) f_{Y|X^*}\left(y|x^*\right) f_{Z|X^*}\left(z|x^*\right) f_{X^*}\left(x^*\right) dx^* \tag{52}$$

In Theorem 8, Assumption (i) weakens independence assumptions used in factor models and can be seen as an analogue of "exclusion restrictions" commonly used in instrumental variables settings. Assumption (ii) was discussed in Section 3.5.1. While the injectivity Assumption (ii) typically requires 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) is a nonclassical centering restriction (see Definition 3).

A general nonlinear and nonseparable relationship (50) can be recovered from the identified densities $f_{Y|X^*}\left(y|x^*\right)$, $f_{X|X^*}\left(x|x^*\right)$, $f_{Z|X^*}\left(z|x^*\right)$ under standard normalizations (e.g. Matzkin (2003)), such as normalizing the distributions of the disturbances

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

to a given distribution. For one-dimensional variables this can be accomplished in the usual way (Cunha, Heckman, and Schennach (2010)):

$$m_V\left(x^*, \Delta v\right) = F_V^{-1}\left(v|x^*\right)$$

if the distribution of ΔV is normalized to a uniform. In multivariate settings, one can rely on recent results on optimal transport and Brenier maps (Carlier, Chernozhukov, and Galichon (2016), Gunsilius (2020)).

The proof of identification of Theorem 8 can be outlined as follows. Assumption (i) directly implies the integral Equation (51). We then associate, for any conditional density $f_{V|U}\left(v|u\right)$, an operator $L_{V|U}$, as in Equation (49). Equation (51) can then be cast as an operator equivalence relationship:

$$L_{y;X|Z} = L_{X|X^*} D_{y;X^*} L_{X^*|Z}, \tag{53}$$

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

$$L_{X^*|Z} = L_{X|X^*}^{-1} L_{X|Z}, \tag{54}$$

substituting it into (53) and rearranging, we obtain:

$$L_{y;X|Z} L_{X|Z}^{-1} = L_{X|X^*} D_{y;X^*} L_{X|X^*}^{-1}, \tag{55}$$

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

Equation (55) 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 $f_{Y|X^*}\left(y|x^*\right)$ and eigenfunctions $f_{X|X^*}\left(\cdot|x^*\right)$ (both indexed by x^*) provide the unobserved densities of interest. To ensure uniqueness of this decomposition, Hu and Schennach (2008) employ four techniques: (i) uniqueness (up to some normalizations) is ensured by Theorem XV 4.5 in Dunford and Schwartz (1971)); (ii) the scale of the eigenfunctions is fixed by the fact that densities integrate to one; (iii) use Assumption (iii) to resolve ambiguities in eigenfunctions if some eigenvalues are degenerates; (iv) invoke Assumption (iv) to uniquely determine the indexing of the eigenvalues and eigenfunctions. The last

step makes use of the centering restriction: Consider another variable $\tilde{x}^*$ related to x^* through $x^* = R(\tilde{x}^*)$, and note that

$$M\left[f_{X|\tilde{X}^*}(\cdot|\tilde{x}^*)\right] = M\left[f_{X|X^*}(\cdot|R(\tilde{x}^*))\right] = R(\tilde{x}^*), \tag{56}$$

which is only equal to $\tilde{x}^*$ if R is the identity function. 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 (54) implies that $f_{X^*|Z}(x^*|z)$ is also identified.

Schennach (2013b) also employs operator diagonalization to cover the converse case of Berkson measurement error in X, where the centering restrictions must be phrased in terms of $f_{X^*|X}$ rather than $f_{X|X^*}$, which requires a substantially different approach. Specifically, the measurement system has the form:

$$\begin{aligned} Y &= g(X^*) + \Delta Y && (57)\\ X^* &= X + \Delta X && (58)\\ Z &= h(X^*) + \Delta Z && (59) \end{aligned}$$

where all three measurement are nonclassical and the functions g and h are unknown.

Theorem 9 *Assumptions that (i) X, ΔX^*, ΔY, ΔZ are mutually independent, (ii) ΔX^*, ΔY, ΔZ are centered (i.e. have zero mean, mode, median, etc.), (iii) $\phi_{\Delta X^*}$ and $\phi_{\Delta Z}$ are nonvanishing, (iv) the functions $g(x^*)$ and $h(x^*)$ are one-to-one, (v) h and its inverse are differentiable. Then, given the true observed conditional density $f_{Y,Z|X}$, the functional equation*

$$f_{Y,Z|X}(y,z|x) = \int f_{\Delta Z}(z - h(x^*)) f_{\Delta Y}(y - g(x^*)) f_{\Delta X^*}(x^* - x)\, dx^* \tag{60}$$

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

Operator diagonalization techniques are now being used to solve an increasing number of identification problems with nonclassical disturbances, such as dynamic models (Hu and Shum (2012), Hu and Shum (2013)), dynamic factor models (Cunha, Heckman, and Schennach (2010)), two-sample combination methods (Carroll, Chen, and Hu (2010)), nonlinear panel data models via quantile restrictions (Arellano and Bonhomme (2016)), with heterogeneity and selection (Sasaki (2015b)), and with interactive fixed effects (Freyberger (2017)). Wilhelm (2015) combines operator inversion techniques with Fourier methods used in quantile models (Schennach (2008)) to provides a nearly

closed form solution to the identification problem of a general nonlinear panel data model. Schennach and Starck (2020) exploit spatial data to re-construct synthetic nonclassical repeated measurements from nearby observations to enable measurement error-robust estimation.

So far, operator diagonalization techniques have been preferentially used as a technique for proving identification rather than as a tool to build explicit estimators. In practice, one typically proceeds by replacing all unknown functions in the appropriate integral equations (such a Equations (52) or (60)) by either parametric forms or nonparametric sieve approximations (Chen (2005)). One then numerically optimizes the unknown coefficients of these expressions, either in a maximum likelihood (Shen (1997)) or GMM framework (Ai and Chen (2003)). Naturally, it would be risky to approach a new latent variable problem by simply writing down an integral equation and attempting to solve it numerically without first obtaining the theoretical guarantee of the existence of a unique solution (for instance using the techniques from this section). The illusion of a unique solution might then simply be the result of the functional form assumptions made either explicitly or implicitly via the truncation of an infinite approximating series.

3.5.3 Discrete variables

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 typically considered a nonclassical error because, when the number of possible values X^* and X can take is finite and if X^* and X have the same number of support points, 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. The case when some of the variables have discrete support (with a finite number of points) presents the unique advantage that one can typically exhaustively write out the finite number of equations relating the observables to the unobservable distributions. This possibility has lead to a number of closed form solutions for specific models (Hu (2008), Chen, Hu, and Lewbel (2008a), Chen, Hu, and Lewbel (2008b), Mahajan (2006), Lewbel (2007), Molinari (2008), Chen, Hu, and Lewbel (2009)). In parametric discrete response models with misclassification, Cameron, Li, Trivedi, and Zimmer (2004), Hausman, Abrevaya, and Scott-Morton (1998) and Li, Trivedi, and Guo (2003) observe that identification without additional information is made possible by the following fact: The errors in the response produce distortions in the observable distribution of the mismeasured response that are distinct from the distributions that could be generated from the underlying parametric model

of the true unobserved response.

The general problem of identifying the distribution of an unobserved discrete variable from misclassified repeated measurements has received considerable attention in the statistics literature (see, e.g., Kruskal (1989), Kolda and Bader (2009) for reviews). One of the most significant results is Kruskal's tensor array decomposition (Kruskal (1977)), which we now describe. Assume that we have access to three potentially misclassified measurements (X, Y, Z) of X^* that are mutually independent conditional on X^*, i.e., $f_{X,Y,Z|X^*}(x,y,z|x^*) = f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*)$, where f with appropriate subscripts denotes a probability mass function. This equality lets us derive a discrete analogue of Equation (52):

$$f_{X,Y,Z}(x,y,z) = \sum_{x^* \in \mathcal{X}} f_{X|X^*}(x|x^*) f_{Y|X^*}(y|x^*) f_{Z|X^*}(z|x^*) f_{X^*}(x^*) \tag{61}$$

where $\mathcal{X}$ denotes the discrete support of the distribution of X^*. We can associate each quantity on the right-hand side with (possibly rectangular) matrices: $A_{ir} \equiv f_{X|X^*}(i|r)$, $B_{jr} \equiv f_{Y|X^*}(j|r)$ and $C_{kr} \equiv f_{Z|X^*}(k|r) f_{X^*}(r) = f_{Z,X^*}(k,r)$, and relabel the support points of each variables as consecutive integral numbers, without loss of generality. Similarly the left-hand side can be associated with a third order tensor (or *three-way array*) $T_{ijk} \equiv f_{X,Y,Z}(i,j,k)$. The question of identification is addressed in the following definition and associated theorem:

Definition 6 *The* Kruskal rank *of a matrix M, denoted K_M, is the largest k such that any k columns of M are not colinear. Note how this definition differs from the usual rank, which is the largest k such that there merely exists one choice of k columns that are noncolinear.*

Theorem 10 *(Kruskal's theorem) If a "three-way array" T admits the decomposition:*

$$T_{ijk} = \sum_{r=1}^{R} A_{ir} B_{jr} C_{kr} \tag{62}$$

with

$$K_A + K_B + K_C \geq 2R + 2, \tag{63}$$

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

$$T_{ijk} = \sum_{r=1}^{R} \tilde{A}_{ir} \tilde{B}_{jr} \tilde{C}_{kr} \tag{64}$$

will satisfy $\tilde{A} = AD_AP$, $\tilde{B} = BD_BP$, $\tilde{C} = CD_CP$ where D_A, D_B, D_C are diagonal matrices satisfying $D_AD_BD_C = I$ and P is a permutation matrix.

The proof of this result is notoriously difficult, although one relatively accessible proof can be found in Rhodes (2010), who starts from the simple case $K_C = 1$ (which is analogous to the matrix diagonalization argument of Hu (2008)) and proceeds by induction to reach the $K_C > 1$ cases.

When the matrices A, B, C represent probabilities, we can use the fact that $\sum_i A_{ir} = \sum_x f_{X|X^*}(x|x^*) = 1$, to uniquely determine D_A. Similarly, $\sum_j B_{jr} = 1$ uniquely determines D_B. Next, D_C is recovered from $D_C = D_A^{-1} D_B^{-1}$ (because $\sum_k C_{kr} \neq 1$ in general, since C_{kr} represents a joint distribution rather than a marginal).

The result effectively states that, under the rank Condition (63), for a given observable $f_{X,Y,Z}(x, y, z)$, Equation (61) admits a unique tuple of unobserved probability mass functions $f_{X|X^*}(x|x^*)$, $f_{Y|X^*}(y|x^*)$, $f_{Z,X^*}(z, x^*)$ as a solution, apart from a trivial re-ordering of the unobservable x^* (via the permutation matrix P). The permutation ambiguity can be lifted by imposing additional natural centering-like conditions (Hu (2008), Hu and Xiao (2018)): For instance, $f_{X|X^*}(x|x^*)$ for a given x^* is maximized at $x = x^*$. In discrete models, the centering restrictions can be considerably weaker, relative to the continuous case, because in the continuous case, it is possible to reparametrize x^* without changing ordering (e.g. $x^* = (\tilde{x}^*)^3$), while in the discrete case, re-orderings are the only possible one-to-one mappings between two random variables with same number of support points.

Kruskal's result is particularly adapted to factor models, because all variables of the model are treated in a symmetric fashion. The Kruskal-based approach to the identification of discrete models is receiving increasing attention in the statistics and econometrics literature (e.g., Allman, Matias, and Rhodes (2009), Bonhomme, Jochmans, and Robin (2016a)). However, since verifying the Kruskal rank of a matrix is a computationally demanding combinatorial problem, there is interest in instead providing conditions under which the model is identifiable with probability one (e.g. Allman, Matias, and Rhodes (2009), Connault (2014)) when one considers models as being drawn at random from a set of possibilities. Although the case of the Kruskal rank is more complex, an analogy with matrix invertibility best illustrates this point: If the elements of a square matrix are drawn independently at random from nondegenerate continuous distributions, there is zero probability that this matrix is not invertible, which arguably eliminates the need for the more computationally demanding check that its determinant is indeed nonzero in a specific application.

It should be noted that identification results for discrete variables *do not* imply identification of the corresponding model with continuous variables via a simple limiting argument. To illustrate this, let $\mathrm{Id}(M)$ denote a function equal "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 $\lim_{n\to\infty}\mathrm{Id}(M_n) = \mathrm{Id}(\lim_{n\to\infty} M_n)$ in general. Therefore, when constructing a sequence of identified discrete models converging to a continuous model, we cannot conclude that the limiting continuous model is identified. A possible way to approach the continuous case would be to instead derive an identified set for the discrete case and show that this set converges to a point (e.g. in the Hausdorff metric) as the discretization becomes finer (as in Williams (2019)).

3.6 Bounding techniques

As we have seen in the previous sections, securing identification of a latent variable model may require strong conditions (such as independence restrictions, potentially after conditioning). When one is unwilling to make such assumptions, one may instead seek to derive constraints on the parameters of interest under weaker assumptions. Such bounding techniques have direct connections to the set identification literature (e.g. Chernozhukov, Hong, and Tamer (2007), Manski (2003), Chandrasekhar, Chernozhukov, Molinari, and Schrimpf (2012), Bontemps, Magnac, and Maurin (2012), Magnac and Maurin (2008), Chesher (2013), Molinari (2018)), in which the goal is to obtain a set Θ of possible values of the parameters consistent with the observed data. In the following sections, we report the bounds in the population but natural sample analogues provide estimates of the bounds.

3.6.1 Positivity

A very natural avenue to derive bounds under minimal assumptions is to exploit the fact that probabilities are positive (and, consequently, that covariance matrices are positive definite). In the linear errors-in-variables model (Equations (45) and (46)), this idea underlies the well-known bounds attributed to Frisch (1934), which is perhaps one of the earliest example of set-identification. These bounds can be derived under the assumption of mutual uncorrelatedness of X^*,ΔX, ΔY and observing that the full covariance structure of X and Y can be expressed in terms of the variances of the unobservable variables $\sigma^2_{X^*}$, $\sigma^2_{\Delta X}$, $\sigma^2_{\Delta Y}$. Imposing the latter two to be positive leads to the inequalities on the slope coefficient θ:

$$\frac{|\mathrm{Cov}\left[X,Y\right]|}{\mathrm{Var}\left[X\right]} \leq |\theta| \leq \frac{\mathrm{Var}\left[Y\right]}{|\mathrm{Cov}\left[X,Y\right]|}$$

(with $\operatorname{sgn}\theta = \operatorname{sgn}\mathrm{Cov}\left[X,Y\right]$). These bounds have the straightforward interpretation as the slope coefficient of the "forward" regression of Y on X (lower bound) and of

the reciprocal of slope coefficient of the "reverse" regression of X on Y (upper bound). These bounds are sharp under uncorrelatedness of X^*,ΔX, ΔY (but not under their mutual independence, see Section 3.4.4).

This result has been formally generalized by Klepper and Leamer (1984) to multivariate linear regression where all regressors $X^* = (X_1^*, \ldots, X_{d_X}^*)'$ are measured with error. It suffices to sequentially use each of the $d_{X^*}+1$ observed variables (Y, X_1,..., X_{d_X}) as the dependent variable and the remaining variables as regressors. After rearrangement, each regression line $k = 1, \ldots, d_{X^*}$ can be cast in the form $Y^* = \sum_{j=1}^{d_{X^*}} \theta_j^{(k)} X_j^*$ (where $Y^* \equiv Y - \Delta Y$), so that the regression coefficient vectors $\theta^{(k)}$ all have the same units. Then, the set obtained by taking the convex hull of $\theta^{(1)}, \ldots \theta^{(d_{X^*})}$ provides the *identified set*, the true set of possible values of the regression coefficients, provided the $\theta^{(k)}$ all lie in the same *orthant*.[3] Otherwise, the set is unbounded along some direction(s). While the above methods sometimes deliver fairly large identified sets, they can often be further narrowed down by imposing a priori plausible restrictions on the variance and on the correlation structure of the measurement error (Klepper and Leamer (1984), Erickson (1993), Hyslop and Imbens (2001)).

Exploiting the positivity of probabilities is also useful in discrete variable models (Black, Berger, and Scott (2000), Bollinger (1996), Molinari (2008), Klepper (1988)) and can often be combined with other plausible constraints (such as bounds on the misclassification probabilities, symmetry, monotonicity constraints, etc.) to yield usefully small identified sets in concrete empirical settings (Molinari (2008), Kreider and Pepper (2008)). More specifically, monotonicity restrictions in themselves often provide effective ways to weaken identifying assumptions while maintaining good set-identification power (Manski and Pepper (2000)).

3.6.2 Interval-valued data

Interval-valued data occur in a number of settings. In surveys, respondents are often asked to report ranges instead of specific values or data may be rounded. Alternatively, some data could be missing or erroneous with some probability (i.e., contaminated or corrupted data), in which case the suspect data could take any value in some prespecified reasonable set.

In such settings, one needs to consider any value of data in the interval as a possibility. If the model parameters are monotone in the input data, the identified set boundaries are determined by the extreme points of the data intervals. Examples of this line of work include Manski and Tamer (2002), who derive bounds on the re-

[3] An orthant is a set of vectors whose elements share the same pattern of signs.

gression 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.

3.6.3 General bounds under nonmonotonicity

The derivation of sharp bounds in general nonlinear latent variables problem is complicated by the potentially nonmonotone relationships among the unobserved variables, the observed variables and the model parameters. For this reason, it is often more useful to rely on general numerical methods to derive the bounds numerically. A number of methods have recently been proposed (Galichon and Henry (2013), Ekeland, Galichon, and Henry (2010), Schennach (2014), Beresteanu, Molchanov, and Molinari (2011) and Bar and Molinari (2017)). We focus here on the method proposed by Schennach (2014) — The reader is referred to the excellent review by Molinari (2018) for a more detailed description of the other methods.

Schennach (2014) solves for the possible value(s) of a parameter vector θ that satisfy a set of moment conditions that are known to hold in the population:

$$E\left[g\left(U,Z,\theta\right)\right]=0 \tag{65}$$

where g is a d_g-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. It is the unobservable U that enables the treatment of measurement systems, with U typically playing the role of X^* and expressing ΔX as a function of U and the observable variables Z. We 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 loss of generality that $\mathcal{U}$ does not depend on z, since this dependence could be incorporated into the function $g\left(u,z,\theta\right)$). Expectations are calculated under the distribution specified as a subscript. The method then relies on the following equivalence:

Theorem 11 *In model (65), for any θ and any distribution π,*

$$\inf_{\mu\ll\rho}\left\|E_{\mu\times\pi}\left[g\left(U,Z,\theta\right)\right]\right\|=0 \tag{66}$$

if and only if

$$\inf_{\gamma\in\mathbb{R}^{d_g}}\left\|E_{\pi}\left[\tilde{g}\left(Z,\theta,\gamma\right)\right]\right\|=0 \tag{67}$$

where

$$\tilde{g}(z,\theta,\gamma) \equiv \frac{\int g(u,z,\theta)\exp(\gamma' g(u,z,\theta))\,d\rho(u|z)}{\int \exp(\gamma' g(u,z,\theta))\,d\rho(u|z)} \tag{68}$$

where ρ is a user-specified dominating[4] *conditional distribution such that (i) $\rho(\cdot|z)$ is supported on $\mathcal{U}$ at each z and (ii) $E_\pi[\ln E_\rho[\exp(\gamma' g(U,Z,\theta))|Z]]$ exists and is twice differentiable in γ for all $\gamma \in \mathbb{R}^{d_g}$ and all θ.*

This theorem turns the original problem of interest (66), which involves optimization over an infinite-dimensional quantity μ, into a *finite* dimensional optimization (67). This simplification is made possible by the fact that the so-called parametric least-favorable entropy maximizing family of distributions (proportional to $\exp(\gamma' g(u,z,\theta))\,d\rho(u|z)$) used to compute the expectation over U in (68) 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 $E[\tilde{g}(z,\theta,\gamma)] = 0$, even in finite samples (since the optimization over γ would yield the same profiled objective function in terms of θ).

Measures $\rho(u|z)$ satisfying the needed 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 (2014).

The fundamental reason Theorem 11 holds is simple. Even though Equation (66) has potentially an infinite number of solutions, we only need to find one. Hence, we can rank distributions according to some criterion (here, their entropy) and convert an "existence" problem into an optimization problem under the constraint (65). This constrained entropy maximization problem has a unique solution which turns out to have the convenient form (68).

A practical way to evaluate the integral (68) is to draw random vectors u_j, $j = 1,\ldots,R$ from a density proportional to $\exp(\gamma' g(u,z,\theta))\,d\rho(u|z)$ using, e.g., the Metropolis algorithm and calculate the average

$$\frac{1}{R}\sum_{j=1}^{R} g(u_j,z,\theta) \tag{69}$$

[4] A measure ρ is said to be dominating μ (written as $\mu \ll \rho$) if μ admits a density with respect to the measure ρ. This is a technical condition that rules out irrelevant events of probability zero.

for a sufficiently large R. 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)). This method, called Entropic Latent Variance Integration via Simulation (ELVIS) can thus be seen as a semiparametric generalization of the MSM.

Averaging over the unobservables then provides a conventional vector of moment conditions $E\left[\tilde{g}\left(Z,\theta,\gamma\right)\right]=0$ involving only observable variables that is equivalent to the original moment condition (65). As a result, solving for the parameter θ of interest and for the nuisance parameter γ can be accomplished through a variety of standard GMM-type techniques (Hansen (1982), Owen (1988), Newey and Smith (2004), Owen (1990), Imbens, Spady, and Johnson (1998), Kitamura and Stutzer (1997), Qin and Lawless (1994), 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.

One should be careful not to interpret the entropy maximizing distribution as the actual distribution of the unobservables. Nevertheless, one can easily bound features of the true distribution of U, provided that these features can be expressed as moment conditions. For instance, if one wished to find all possible values of the mean of U consistent with the data, one could merely include $E\left[U-\tilde{\theta}\right]=0$ among the moment conditions and the identified set for the auxiliary parameter $\tilde{\theta}$ will indicate the possible values of $E\left[U\right]$. If one wished to bound the cdf of U at some point u, the appropriate moment condition would be $E\left[\mathbf{1}\left\{U\leq u\right\}-\tilde{\theta}\right]$. The key is that the least favorable family will automatically adapt to the addition moment conditions to explore the possible worst case scenarios for the moment of interest.

The ELVIS method is useful to estimate the identified set in general nonlinear regression models with covariate measurement error under very weak uncorrelatedness assumptions when no additional indicators or instruments are available, as shown in Schennach (2014). This work also shows that ELVIS nests as special case both the higher-order moment treatment and the bounding treatment of the standard linear errors-in-variables model. A linear regression with measurement error estimated using higher order moments results in a sharply peaked objective function for non-normal data, reflecting the point identification result of Reiersol (1950) in that case. But for normal data, the objective function becomes maximized over an interval that reveals

the identified set obtained from the forward and reverse regression bounds of Section 3.6.1.

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 (2014) provides extensions to conditional mean and independence restrictions. Recently, the fact that ELVIS reaches the semiparametric efficiency bound has been formally established by Bedard and Renault (2016) for point-identified models.

The ELVIS method is related to moment inequality methods (e.g. Beresteanu, Molchanov, and Molinari (2011)), as shown by the following result.

Theorem 12 *The identified set Θ_0 can be equivalently described by*

$$\{\theta \in \Theta : E_{\pi_0}[t(Z,\theta,\eta)] \geq 0 \text{ for all } \eta \in \delta\mathcal{B}_1\},$$

where $\delta\mathcal{B}_1 = \{\eta \in \mathbb{R}^{d_g} : \|\eta\| = 1\}$ (the unit ball boundary) and

$$t(Z,\theta,\eta) \equiv \lim_{r\to\infty} \eta' \tilde{g}(Z,\theta,\eta r) \tag{70}$$

for $\tilde{g}(Z,\theta,\gamma)$ as in Theorem 11. Note that if, for some η, the limit in (70) diverges then no constraint is associated with this value of η. An alternative expression is

$$t(Z,\theta,\eta) = \sup_{u\in\mathcal{U}} \eta' g(u,Z,\theta). \tag{71}$$

This result shows that a general GMM model with unobservables can be expressed as an infinite set of moment inequalities (indexed by η). A feasible way to implement this alternative approach is to re-write the problem as an optimization over η that seeks the tightest inequality (as done in Beresteanu, Molchanov, and Molinari (2011) under the assumption that $t(Z,\theta,\eta)$ is bounded). However, this optimization problem is, in general, nonsmooth, whereas the finite-dimensional ELVIS optimization problem of Theorem 11 is smooth by construction.

4 Application examples

The techniques outlined so far are making a significant impact on empirical economic work. Below, we delve deeper into a number of illustrative examples picked from a range of diverse fields. Our list is by no mean exhaustive and the works are selected in large

part on the basis of how they demonstrate the practical use of the different methods reviewed here. A common feature of all of these applications is that they do not rely on the availability of an auxiliary sample of perfectly measured validation data, nor do they merely assume that the measurement error has a known variance or distribution. Instead, these studies exploit the availability of multiple measurements or indicators that, together, enable identification of the feature of the unobservables. The success and broad applicability of this approach suggests that future data collection efforts should be designed with this in mind, as noted by other researchers (e.g., Browning and Crossley (2009)).

4.1 Human capital

Measurement of human capital (i.e. skills or abilities) has historically been one of the earliest uses of factor models, since the well-known Spearman (1904) "general intelligence" factor was proposed. This remains, to this day, a very frequent application of factor models (e.g. Cunha, Heckman, and Schennach (2010), Cunha and Heckman (2008), Attanasio, Meghir, and Nix (2020), Attanasio, Cattan, Fitzsimons, Meghir, and Rubio-Codina (2019)). The basic idea is that any single measure of skills is bound to be highly noisy and simple averages of these measures do not succeed in fully averaging out the noise. However, by accounting for the measurement error in a factor model framework, one can make robust inference regarding statistical features of the skills even if each individual's skill level is not known precisely.

A popular topic of investigation is the dynamics of skill acquisition over time in response to parental or governmental inputs. Cunha, Heckman, and Schennach (2010) set up a "production function" (or "technology") that relates future skills X_{t+1}^* to current skills X_t^*, various inputs I and covariates C:

$$X_{t+1}^* = g\left(X_t^*, I, C, U\right)$$

where U is a disturbance. The unobserved skills X_t^* are assume to be related to observables via a measurement system: $X_t = m\left(X_t^*, \Delta X_T\right)$ (assumed to be linear in their application, although it is not required by their general theoretical approach). A key feature of their approach is to allow for nonlinear production functions g so that substitution effects can be investigated. The development of such structural model offers opportunities to carry out counterfactual analyses (Low and Meghir (2017)) that investigate different re-allocations of the investments across children of different characteristics or across time periods (Cunha, Heckman, and Schennach (2010), Attanasio, Meghir, and Nix (2020)). The application of measurement systems in this

field goes beyond the measurement of evolving cognitive skills to also include imperfect measurement of parents skills, children's noncognitive skills, parental investments into developing their children's skills (Cunha, Heckman, and Schennach (2010), Cunha and Heckman (2008)), as well as children's health (Attanasio, Meghir, and Nix (2020)).

4.2 Measuring economic growth

It is generally recognized that officially reported GDP for developing countries may not be very accurate and Hu and Yao (2018) nicely illustrate the idea of combining multiple imperfect nonclassical measurements to address this. They consider the measurement of true unobserved per capita GDP (X^*) using (i) the World Bank's reported GDPs (X), (ii) satellite measurement of night time light intensity (Z) and (iii) two other auxiliary variables, a country's latitude (L) and its so-called "statistical capacity" (S), which measures the data collecting abilities of the country's statistical institutions.

Under suitable conditional independence assumptions, they express the joint density of the observables as:

$$f(z, x, s, l) = \int f(z|x^*, l) f(x|x^*, s) f(x^*, s, l) dy^*$$

and employ an operator diagonalization argument to show that the observed left-hand side density uniquely determines the unobserved right-hand side densities. This result also relies on the assumption that reported GDP exhibits zero mean error for one value s_0 of statistical capacity S. Reported GDP for $S \neq s_0$ are allowed to be nonclassical and so are all the other indicators of GDP. This setup does not exactly fall within the Hu and Schennach (2008) framework, due to the presence of the additional variables S and L (which cannot be eliminated by mere conditioning), but operator diagonalization techniques can be suitably adapted nevertheless. For estimation purposes, all unknown densities are represented by sieve expansions.

This setup enables them to uncover the distribution of measurement errors on z and x and the potentially nonlinear relationship between light intensity and true GDP. This helps corroborate some features of this data that have been postulated, but not yet firmly empirically established.

At low levels of GDP, light intensity shows a strong correlation with real GDP, presumably because at early stages of development, most development efforts are aimed at building basic infrastructure. In contrast, at high levels of GDP, light intensity shows little dependence on real GDP, suggesting that economies turn their focus to technological innovations. In addition, the magnitude of the measurement error on reported

GDP is clearly larger for developing economies (with low GDP and low statistical capacity). These features call for methods that can handle nonclassical measurements.

As discussed in the paper, this framework also provides a second look at the controversial issue of China's actual GDP growth, suggesting that its GDP growth may indeed have been overreported in the official figures, although not by as much as estimates based solely on light output would have suggested.

4.3 Income and consumption dynamics

Models of income and consumption behavior often rely on unobserved variables to go beyond simple static exercises and "representative individual" arguments. For instance, Arellano, Blundell, and Bonhomme (2017) investigate income and consumption dynamics emphasizing the importance of individual heterogeneity and nonlinearity.

They develop a model of the time evolution of income as well as of consumption and assets accumulation decisions. Latent variables enter the model at multiple levels and their properties can be inferred using a measurement system framework. First, earning dynamics is governed by a nonlinear hidden Markov process, where the observed income X_t at time t is the sum of a permanent component X_t^* and a transitory component ΔX_t:

$$X_t = X_t^* + \Delta X_t. \tag{72}$$

(we omit individual index i for conciseness). The transitory component plays the role of a white noise measurement error (independent from X_t^* over all times), while the permanent component is the latent variable of interest, which is assumed to follow a general first order Markov process:

$$X_{t+1}^* = g\left(X_t^*, \varepsilon_t\right),$$

where ε_t is a sequence of independent and identically distributed disturbances. It is this Markov structure that enables the use of consecutive observations of income to play the role of multiple indicators of the permanent component X_t^* at time t. They then make use of the hybrid operator and Kotlarski-type approach of Wilhelm (2015), to show identification of the joint distribution all the unobserved variables $X_t^*, \Delta X_t, \varepsilon_t$, under suitable conditional independence assumptions.

The second role played by latent variables is in the consumption dynamics equation, where individual heterogeneity (e.g. tastes) influences the consumption level of the household. In this part of the model, the observed variable is consumption in three consecutive time periods $(\tilde{X}_t, \tilde{X}_{t+1}, \tilde{X}_{t+2})$, whose distribution depends on (i) an unobserved time-invariant taste $\tilde{X}^*$, (ii) other observable covariates (such as assets and

earnings) and (iii) other unobserved variables whose distributions were identified in the previous step (i.e., the permanent component of earnings). Upon implicitly conditioning on those known quantities, the identification problem reduces to solving the integral equation of the form:

$$f\left(\tilde{x}_t, \tilde{x}_{t+1}, \tilde{x}_{t+2}\right) = \int f\left(x_t|\tilde{x}^*\right) f\left(x_{t+1}|\tilde{x}^*\right) f\left(x_{t+2}|\tilde{x}^*\right) f\left(\tilde{x}^*\right) d\tilde{x}^*,$$

which again falls within the domain of applicability of the operator diagonalization methods (discussed in Section 3.5.2).

The above summary of this work somewhat understates its true complexity. Most of the functions actually depend on a number of other variables and the complete dependence structure of these variables over time must be carefully specified and accounted for. Estimation also presents challenges, given the large number of flexibly specified functional forms for all the distributions and nonlinear functions employed.

4.4 Game theory

In game-theoretic settings, the players' actions may depend on unobserved variables (Hu and Shum (2013), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011)), either because payoffs are not fully observable to the econometrician or because there are multiple possible equilibria and the econometrician does not know which one the players are selecting.

A more specific example is the one of auction models (Li, Perrigne, and Vuong (2000), Athey and Haile (2002), Krasnokutskaya (2011), An, Hu, and Shum (2010), Guerre, Perrigne, and Vuong (2000)), where the bids made by different players are related to their valuation, which may consist of both a common and a private component. There is a close analogy between these setups and a measurement system in which the bids play the role of multiple measurements of the same common value that are each altered in different ways by each player's private value. An important step in these approaches is to first exploit the structure of the auction to obtain an explicit relationship between the bids and the valuations (see Athey and Haile (2007) for a review). A classical example can be found in Li, Perrigne, and Vuong (2000), where they exploit the fact that the observed bids b_j for strategic players $j = 1, \ldots, J$ implicitly indicate their valuations, shifted by an amount $F_{B|b}\left(b_j|b_j\right) / f_{B|b}\left(b_j|b_j\right)$ where $B_j = \max_{j' \neq j} b_{j'}$. After assuming a classical error structure in log levels and a symmetric auction, they obtain a measurement system:

$$X_j = X^* + \Delta X_j$$

where X^* represents the log common value and ΔX_j is the log private value of player j and X_j plays the role of a measurement, obtained from the bids via:

$$X_j \equiv \ln\left(b_j + \frac{F_{B|b}\left(b_j|b_j\right)}{f_{B|b}\left(b_j|b_j\right)}\right).$$

Implementing this method of course requires the nonparametric estimation of the appropriate conditional densities and cdf, which fortunately only involve the observable bids.

Latent variable techniques also prove useful to handle other issues pertaining to auctions. For instance, An, Hu, and Shum (2010) address the issue that the number of potential bidders (which has an effect on the player's optimal strategy) is generally unobserved. They exploit the availability of various indicators (e.g., other bids, an observed maximum possible number of auctioneers and other variables that are known to be affect number of bidders, such as the time of day) to identify the model. They account for the fact that the bidders' optimal strategy may depend on the number of bidders and use an instrument that affect the number of bidders entering an auction but that satisfies the exclusion restriction that they have otherwise no direct effect on the bidders' strategy.

4.5 Revealed preferences under heterogeneity and measurement error

While revealed preferences arguments (e.g., McFadden (2005), Afriat (1973), Varian (1982)) are widely accepted and well understood, their empirical use is complicated by the fact that heterogeneity in consumer preferences and their individual-specific discount factors are not directly observed, and by the fact that prices and/or quantities may be subject to measurement error. Allowing for these possibilities is important in order to avoid spurious empirical rejections of the basic revealed preferences assumptions, but also leads to more intricate latent variable models that need to be handled via advanced techniques.

Recently, researchers have begun to tackle these issues in substantial generality by taking advantage of latent variables techniques. For instance, Illanes (2015) models pension plan selection in the presence of switching costs when consumers exhibit unobserved preference heterogeneities. The latter cannot simply be dealt with by attempting to average out the heterogeneity, due to both endogeneity in the decisions, and the fact that averaging preference inequalities leads to a weaker constraint than imposing them at the individual level.

Illanes casts the revealed preference condition as a moment condition of the general form:

$$E\left[\mathbf{1}\left\{s=j\right\}\sum_{j'\in\mathcal{J}\setminus\{j\}}\mathbf{1}\left\{u_{j'}+\Delta\geq u_j+\zeta_{jj'}\right\}\right]=0 \text{ for } j\in\mathcal{J} \tag{73}$$

(for each consumer at each time period), where s is the choice the consumer makes out of the set $\mathcal{J}$, while u_j represents the total present and future expected discounted utility if choice j is made, Δ are the switching costs (if the choice j in consecutive periods differ) and $\zeta_{jj'}$ accounts for unobserved individual preferences. In practice, the u_j and Δ terms are nontrivial to obtain, as u_j is the solution to a stochastic dynamic programming problem (accounting for all future costs and returns) and Δ depends on the history of choices. To ensure identification of the parameters of interest (that enter u_j and Δ), he makes use of the fact, (shown in Chesher, Rosen, and Smolinski (2013)) that a multinomial discrete choice model is identified using a revealed preference moment (such as Equation (73)) and an independence restriction between an observed instrument Z and the unobservable $\zeta_{jj'}$. He implements this constraint as a truncated sequence of moment conditions: $E\left[\left(Z-E\left[Z\right]\right)\left(\zeta_{jj'}\right)^k\right]=0$ for $k=1,2$. The fact that the moment conditions involve unobservables is handled using the general ELVIS framework of Schennach (2014). This avoids the need to devise a complex estimator that explicitly solves for the parameter of interest in terms of the observable using a generic simulation-based approach. This application demonstrates the use of instruments in a rather new context where the exclusion restriction involves unobservables, so that existing instrumental variable estimators are inapplicable. It also shows how devising latent variable models with minimal assumptions enables the determination of empirically relevant bounds on parameters of economic interest (in this application, a lower bound on the costs of switching pension plans).

The ELVIS framework has also been used by Aguiar and Kashaev (2020) to devise more realistic tests of the fundamental exponential discounting model within a revealed preferences framework. The idea is that the consumers' exponentially discounting behavior can be masked if there is heterogeneity in their discount factors as well as measurement errors in prices and/or quantities. The ELVIS is ideally suited to handle such questions as it makes it possible to directly incorporate the discount factor and the measurement error as latent variables and cast the revealed preference constraints as moment conditions. This work goes beyond a simple application of ELVIS, however: the authors also implement, without parametric restrictions, a concavity constraint on the consumer's utilities as a set of moment inequalities, which allows them to account for infinite dimensional unobservables in a feasible fashion. The resulting

moment inequalities can also be incorporated in the ELVIS objective function by introducing appropriate slackness parameters. Finally, they devise ways to implement many economically-justified forms of nonclassical measurement errors in the form of moment conditions. The end result of this effort is perhaps one of the first fully nonparametric test of the exponential discounting hypothesis.

Their empirical findings are informative: They find support for exponential discounting behavior for single individuals (even in previously studied datasets why simpler models had rejected the hypothesis). However, they reject the exponential discounting behavior hypothesis for couples. This suggests that, even though very general and sophisticated models always carries the risk of increasing standard errors to the point where rejections are rare, their approach still exhibit good power against alternatives and can thus meaningfully discriminate between data generating processes that do and that do not conform to the exponential discounting hypothesis.

5 Conclusion

This review gives a snapshot of some of the most powerful techniques currently available to establish the identification of various features (e.g. moments, distributions) of unobserved variables based on a measurement system relating all observed and unobserved variables in very general ways.

The available techniques cover a quite broad range of mathematical techniques, including Fourier transforms, higher-order moments, operator theory, tensors, nonlinear optimization, etc. A notable feature of the existing results is that, often, even very small changes in the model's structure or in its assumption demands completely different approaches. This suggests the possibility of devising more fundamental and general approaches that encompass some of these currently disparate cases within a common framework. Another observation is that some of the most powerful techniques (in terms of range of applicability) do not necessarily provide constructive identification results that can be used for estimation. Conversely, results that may be the most convenient for estimation purposes may not deliver transparent conditions for point identification. This situation points to an active area for future research.

We have seen that numerous empirical applications of the techniques described herein are emerging and empirical researchers are increasingly willing to incorporate latent variable techniques in their toolkit. The fact that most economic data is imperfect in some way does not need to be viewed as mostly an annoyance but instead as an opportunity to devise better ways to explain the available observable data. The idea that, figuratively, we cannot directly get at truth by asking a question once but

rather indirectly by "interrogating" the data in different ways is starting to become an intrinsic part of the way economists approach data.

References

Adams, C. P. (2016): "Finite mixture models with one exclusion restriction," *The Econometrics Journal*.

Adusumilli, K., and T. Otsu (2015): "Nonparametric instrumental regression with errors in variables," Discussion paper, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Adusumilli, K., T. Otsu, and Y.-J. Whang (2017): "Inference on distribution functions under Measurement error," Working Paper, London School of Economics.

Afriat, S. N. (1973): "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method," *International Economic Review*, 14, 460–472.

Aguiar, V. H., and N. Kashaev (2020): "Stochastic Dynamic Revealed Preferences with measurement error," *Review of Economic Studies*, forthcoming.

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

Aigner, D. J., C. Hsiao, A. Kapteyn, and T. Wansbeek (1986): "Latent Variable Models in Econometrics," in *Handbook of Econometrics*, vol. II, pp. 1323–1386. Elsevier Science.

Allman, E. S., C. Matias, and J. A. Rhodes (2009): "Identifiability of parameters in latent structure Models with many observed variables," *Annals of Statistics*, 37, 3099–3132.

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.

Andrews, D. W. K. (2017): "Examples of L^2-Complete and Boundedly-Complete Distributions," *Journal of Econometrics*, 199, 213–220.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and consumption dynamics: a nonlinear panel data framework," *Econometrica*, 85, 693–734.

Arellano, M., and S. Bonhomme (2011): "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *Review of Economic Studies*, 79, 987–1020.

Arellano, M., and S. Bonhomme (2016): "Nonlinear panel data estimation via quantile regressions," *Econometrics Journal*, 19, C61–C94.

Arellano, M., and S. Bonhomme (2020): "Recovering Latent Variables by Matching," Working Paper, University of Chicago.

Athey, S., and P. A. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70, 2107–2140.

Athey, S., and P. A. Haile (2007): "Nonparametric approaches to auctions," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. E. Leamer, vol. 6, Part A, pp. 3847–3965. Elsevier Science.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2019): "Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia," Working Paper 1987R3, Cowles Foundation.

Attanasio, O., C. Meghir, and E. Nix (2020): "Human Capital Development and Parental Investment In India," *Review of Economic Studies*, 87, 2511–2541.

Bar, H., and F. Molinari (2017): "Computational Methods for Partially Identified Models via Data Augmentation and Support Vector Machines," Working Paper, Cornell.

Battistin, E., and A. Chesher (2014): "Treatment Effect Estimation with Covariate Measurement Error," *Journal of Econometrics*, 178(2), 707 – 715.

Bedard, M., and E. Renault (2016): "Asymptotic Efficiency in Estimation with Moment Restrictions Including Latent Variables," Working Paper, Brown University.

Ben-Moshe, D. (2018): "Identification of joint distributions in dependent factor models," *Econometric Theory*, 34, 134–165.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79, 1785–1821.

Berkson, J. (1950): "Are there two regressions?," *Journal of the American Statistical Association*, 45, 164–180.

Black, D. A., M. C. Berger, and F. A. Scott (2000): "Bounding parameter estimates with nonclassical measurement error," *Journal of the American Statistical Association*, 95(451), 739–748.

Blundell, R., J. Horowitz, and M. Parey (2018): "Estimation of a Nonseparable Heterogeneous Demand Function with Shape Restrictions and Berkson Errors," Working Paper CWP67/18, Cemmap.

Bollinger, C. R. (1996): "Bounding mean regressions when a binary regressor is mismeasured," *Journal of Econometrics*, 73(2), 387–399.

Bollinger, C. R. (1998): "Measurement Error in the Current Population Survey: A Nonparametric Look," *Journal of Labor Economics*, 16, 576–594.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016a): "Estimating Multivariate Latent-Structure Models," *Annals of Statistics*, 44(2), 540–563.

Bonhomme, S., K. Jochmans, and J.-M. Robin (2016b): "Non-parametric estimation of finite mixtures from repeated measurements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1), 211–229.

Bonhomme, S., and J.-M. Robin (2009): "Consistent Noisy Independent Component Analysis," *Journal of Econometrics*, 149, 12–25.

Bonhomme, S., and J.-M. Robin (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *Review of Economic Studies*, 77, 491–533.

Bontemps, C., T. Magnac, and E. Maurin (2012): "Set Identified Linear Models," *Econometrica*, 80, 1129–1155.

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.

Brainard, W. C., and J. Tobin (1968): "Pitfalls in Financial Model Building," *American Economic Review*, 58, 99–122.

Browning, M., and T. F. Crossley (2009): "Are two cheap, noisy measures better than one expensive, accurate one?," *American Economic Review: Papers and Proceedings*, 99, 99–103.

Cameron, A., T. Li, P. Trivedi, and D. Zimmer (2004): "Modeling the Differences in Counted Outcomes using Bivariate Copula Models with Application to Mismeasured Counts," *Econometrics Journal*, 7, 566–584.

Carlier, G., V. Chernozhukov, and A. Galichon (2016): "Vector quantile regression: an optimal transport approach," *Annals of Statistics*, 44, 1165–1192.

Carrasco, M., and J.-P. Florens (2011): "Spectral method for deconvolving a density," *Econometric Theory*, 27, 546–581.

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., X. Chen, and Y. Hu (2010): "Identification and estimation of nonlinear models using two samples with nonclassical measurement errors," *Journal of Nonparametric Statistics*, 22, 379–399.

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.

Carroll, R. J., D. Ruppert, C. M. Crainiceanu, T. D. Tosteson, and M. R. Karagas (2004): "Nonlinear and Nonparametric Regression and Instrumental Variables," *Journal of the American Statistical Association*, 99, 736–750.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006): *Measurement Error in Nonlinear Models*. New York: Chapman & Hall.

Chandrasekhar, A., V. Chernozhukov, F. Molinari, and P. Schrimpf (2012): "Inference for best linear approximations to set identified functions," Working Papers CWP43/12, Cemmap.

Chen, X. (2005): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, vol. Vol. 6. Elsevier Science.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49, 901–937.

Chen, X., H. Hong, and E. Tamer (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343–366.

Chen, X., H. Hong, and A. Tarozzi (2008): "Semiparametric Efficiency in GMM Models with Nonclassical Measurement Error," *Annals of Statistics*, 36, 489–1030.

Chen, X., Y. Hu, and A. Lewbel (2008a): "Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments," *Economics Letters*, 100, 381–384.

Chen, X., Y. Hu, and A. Lewbel (2008b): "A note on the closed-form identification of regression models with a mismeasured binary regressor," *Statistics and Probability Letters*, 78, 1473–1479.

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.

Chen, X., O. Linton, and Y. Yi (2017): "Semiparametric Identification of the Bid-Ask Spread in Extended Roll Models," *Journal of Econometrics*, 200, 312–325.

Cheng, C.-L., and J. W. V. Ness (1999): *Statistical Regression with Measurement error*. Arnold, London.

Cheng, C.-L., and H. Schneeweiss (1998): "Polynomial Regression with Errors in the Variables," *J. R. Statist. Soc. B*, 60, Part 1, 189–199.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets In Econometric Models," *Econometrica*, 75, 1243–1284.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007): "Instrumental Variable Identification and Estimation of Nonseparable Models via Quantile Conditions," *Journal of Econometrics*, 139, 4–14.

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

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

Chesher, A. (2013): "Semiparametric structural models of binary response: shape restrictions and partial identification," *Econometric Theory*, 29(02), 231–266.

Chesher, A. (2017): "Understanding the effect of measurement error on quantile regressions," *Journal of Econometrics*, pp. 223–237.

Chesher, A., A. M. Rosen, and K. Smolinski (2013): "An instrumental variable model of multiple discrete choice," *Quantitative Economics*, 4, 157–196.

Chesher, A., and C. Schluter (2002): "Welfare Measurement and Measurement Error," *Review of Economic Studies*, 69, 357–378.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Compiani, G., and Y. Kitamura (2016): "Using mixtures in econometric models: a brief review and some new results," *The Econometrics Journal*, 19, C95–C127.

Connault, B. (2014): "Hidden Rust Models," Discussion paper, Princeton Job Market Paper.

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, 883–931.

Cunha, F., and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43, 738–782.

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 (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

DELAIGLE, A. (2007): "Nonparametric density estimation from data with a mixture of Berkson and classical errors," *Canadian Journal of Statistics*, 35, 89–104.

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.

DELAIGLE, A., AND A. MEISTER (2007): "Nonparametric regression estimation in the heteroscedastic errors-in-variables problem," *Journal of the American Statistical Association*, 102, 1416–1426.

D'HAULTFOEUILLE, X. (20011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460–471.

D'HAULTFOEUILLE, X., AND P. FÉVRIER (2015): "Identification of mixture models using support variations," *Journal of Econometrics*, 189, 70–82.

DONG, H., T. OTSU, AND L. TAYLOR (2020): "Estimation Of Varying Coefficient Models with Measurement Error," *Journal of Econometrics*, forthcoming.

DUNFORD, N., AND J. T. SCHWARTZ (1971): *Linear Operators*. John Wiley & Sons, NY.

EBRAHIMI, N., G. HAMEDANI, E. S. SOOFI, AND H. VOLKMER (2010): "A class of models for uncorrelated random variables," *Journal of Multivariate Analysis*, 101, 1859–1871.

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., C. H. JIANG, AND T. M. WHITED (2014): "Minimum distance estimation of the errors-in-variables model using linear cumulant equations," *Journal of Econometrics*, 183.

ERICKSON, T., AND T. M. WHITED (2000): "Measurement Error and the Relationship between Investment and "q"," *Journal of Political Economy*, 108, 1027–1057.

ERICKSON, T., AND T. M. WHITED (2002): "Two-step GMM estimation of the errors-in-variables model using high-order moments," *Econometric Theory*, 18, 776–799.

ERICKSON, T., AND T. M. WHITED (2012): "Treating measurement error in Tobin's q," *Review of Financial Studies*, 25(4), 1286–1329.

EVDOKIMOV, K. (2009): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Working Paper, Yale University.

EVDOKIMOV, K., AND H. WHITE (2012): "Some Extensions of a Lemma of Kotlarski," *Econometric Theory*, 28, 925–932.

EVDOKIMOV, K., AND A. ZELENEEV (2020a): "Errors-In-Variables in Large Nonlinear Panel and Network Models," Working Paper, Cemmap.

EVDOKIMOV, K., AND A. ZELENEEV (2020b): "Simple Estimation of Semiparametric Models with Measurement Errors," Working Paper, Cemmap.

FAN, J. (1991a): "Asymptotic Normality for Deconvolution Kernel Density Estimators," *Sankhyā: The Indian Journal of Statistics*, 53, Series A, Pt. 1, 97–110.

FAN, J. (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.

FREYBERGER, J. (2017): "Nonparametric panel data models with interactive fixed effects," *Review of Economic Studies*, forthcoming.

FRISCH, R. (1934): *Statistical Confluence Analysis by Means of Complete Regression Systems*. University Institute of Economics, Oslo, Norway.

FULLER, W. A. (1987): *Measurement Error models*. Wiley, New York.

Galichon, A., and M. Henry (2013): "Dilation Bootstrap: A methodology for constructing confidence regions with partially identified models," *Journal of Econometrics*, 177, 109–115.

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.

Gospodinov, N., I. Komunjer, and S. Ng (2017): "Simulated minimum distance estimation of dynamic models with errors-in-variables," *Journal of Econometrics*, 200, 181–193.

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.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, 525–574.

Gunsilius, F., and S. M. Schennach (2019): "Independent Principal Component Analysis," Working Paper CWP46/19, Cemmap.

Gunsilius, F. F. (2020): "A condition for the identification of multivariate models with binary instruments," Working Paper, University of Michigan.

Hall, P., and J. L. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904–2929.

Hall, P., and X.-H. Zhou (2003): "Nonparametric estimation of component distributions in a multivariate mixture," *Annals of Statistics*, 31, 201–224.

Hamedani, G. (2013): "Sub-Independence: An Expository Perspective," *Commun. Statist. Theory-Methods*, 42, 3608–3613.

Hamedani, G. G., and H. W. Volkmer (2009): "Comment on 'Conditional Moments and Independence' by A. De Paula.," *The American Statistician*, 63, 295–295.

HANSEN, L. P. (1982): "Large sample properties of generalized method of moment estimators," *Econometrica*, 50, 1029–1054.

HÄRDLE, W., AND O. LINTON (1994): "Applied Nonparametric Methods," in *Handbook of Econometrics*, ed. by R. Engle, and D. McFadden, vol. IV. Elsevier Science.

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., J. ABREVAYA, AND F. M. SCOTT-MORTON (1998): "Misclassification of the Dependent Variable in a Discrete-Response Setting," *Journal of Econometrics*, 87, 237–269.

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.

HAUSMAN, J. A., AND W. E. TAYLOR (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398.

HECKMAN, J. J., AND E. VYTLACIL (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

HENRY, M., Y. KITAMURA, AND B. SALANIÉ (2014): "Partial identification of finite mixtures in econometric models," *Quantitative Economics*, 5, 123–44.

HODERLEIN, S., B. SIFLINGER, AND J. WINTER (2015): "Identification of structural models in the presence of measurement error due to rounding in survey responses," Working Paper, Boston College.

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. (2015): "Microeconomic models with latent variables: Applications of measurement error models in empirical industrial organization and labor economics," Working Paper CWP03/15, Cemmap.

Hu, Y., and G. Ridder (2010): "On Deconvolution as a First Stage Nonparametric Estimator," *Econometric Reviews*, 29, 1–32.

Hu, Y., and G. Ridder (2012): "Estimation of Nonlinear Models with Measurement Error Using Marginal Information," *Journal of Applied Econometrics*, 27, 347–85.

Hu, Y., and Y. Sasaki (2015): "Closed-form estimation of nonparametric models with non-classical measurement errors," *Journal of Econometrics*, 185, 392–408.

Hu, Y., and S. M. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76, 195–216.

Hu, Y., S. M. Schennach, and J.-L. Shiu (2017): "Injectivity of a Class of Integral Operators with Compactly Supported Kernels," *Journal of Econometrics*, 200, 48–58.

Hu, Y., S. M. Schennach, and J.-L. Shiu (2020): "Identification of Nonparametric Regression Models with Nonclassical Measurement Errors," *Journal of Econometrics*, forthcoming.

Hu, Y., and J.-L. Shiu (2013): "Identification and estimation of nonlinear dynamic panel data models with unobserved covariates," *Journal of Econometrics*, 175, 116–131.

Hu, Y., and M. Shum (2012): "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171, 32–44.

Hu, Y., and M. Shum (2013): "Identifying dynamic games with serially-correlated unobservables," *Advances in Econometrics*, 31, 97–113.

Hu, Y., and R. Xiao (2018): "Global estimation of finite mixture and misclassication models with an application to multiple equilibria," Discussion Paper CWP32/18, CeMMAP.

Hu, Y., and J. Yao (2018): "Illuminating Economic Growth," 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.

Illanes, G. (2015): "Switching Costs in Pension Plan Choice," Working Paper, MIT.

Imbens, G. W., R. H. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Jochmans, K., M. Henry, and B. Salanié (2017): "Inference on two-component mixtures under tail restrictions," *Econometric Theory*, 33, 610–635.

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. (2003): "Nonparametric identifiability of finite mixtures," Working Paper, Yale University.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moment Estimation," *Econometrica*, 65, 861–874.

Klepper, S. (1988): "Bounding the effects of measurement error in regressions involving dichotomous variables," *Journal of Econometrics*, 37(3), 343–359.

Klepper, S., and E. E. Leamer (1984): "Consistent Sets of Estimates for Regressions with Errors in all Variables," *Econometrica*, 52, 163–183.

Kolda, T. G., and B. W. Bader (2009): "Tensor Decompositions and Applications," *SIAM Review*, 51, 455–500.

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.

KREIDER, B., AND J. PEPPER (2008): "Inferring disability status from corrupt data," *Journal of Applied Econometrics*, 23, 329–349.

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.

KRUSKAL, J. B. (1989): "Rank, decomposition, and uniqueness for 3-way and n-way arrays," in *Multiway Data Analysis*, ed. by R. Coppi, and S. Bolasco. Elsevier Science Publishers B.V., North-Holland.

LEE, L.-F., AND J. H. SEPANSKI (1995): "Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data," *Journal of the American Statistical Association*, 90, 130–140.

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.

LEWBEL, A. (2007): "Estimation of Average Treatment Effects With Misclassification," *Econometrica*, 75, 537–551.

LEWBEL, A. (2012): "Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics*, 30, 67–80.

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., P. TRIVEDI, AND J. GUO (2003): "Modeling Response Bias in Count: A Structural Approach With an Application to the National Crime Victimization Survey Data," *Sociological Methods and Research*, 31, 514–544.

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.

Low, H., and C. Meghir (2017): "The Use of Structural Models in Econometrics," *Journal of Economic Perspectives*, 31, 33–58.

Magnac, T., and E. Maurin (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75, 835–864.

Mahajan, A. (2006): "Identification and Estimation of Single Index Models with Misclassified Regressor," *Econometrica*, 74, 631–665.

Mallick, B., F. O. Hoffman, and R. J. Carroll (2002): "Semiparametric Regression Modeling with Mixtures of Berkson and Classical Error, with Application to Fallout from the Nevada Test Site," *Biometrics*, 58, 13–20.

Mallows, C. (2007): "Deconvolution by Simulation," in *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, ed. by R. Liu, W. Strawderman, and C. Zhang. Institute of Mathematical Statistics, Beachwood, Ohio, USA.

Manski, C. (2003): *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

Manski, C. F., and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010.

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.

MATZKIN, R. L. (2008): "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76, 945–978.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

MCFADDEN, D. L. (2005): "Revealed stochastic preference: a synthesis," *Economic Theory*, 26, 245–264.

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.

MOLINARI, F. (2018): "Econometrics with Partial Identification," in *Handbook of Econometrics*, vol. 7A, forthcoming. Elsevier Science.

NADAI, M. D., AND A. LEWBEL (2016): "Nonparametric Errors in Variables Models with Measurement Errors on both sides of the Equation," *Journal of Econometrics*, 191, 19–32.

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

NEWEY, W., AND R. J. SMITH (2004): "Higher-Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

OTSU, T., AND L. TAYLOR (2016): "Specification Testing for Errors in Variables models," Discussion paper, LSE.

OWEN, A. B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

OWEN, A. B. (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.

Rhodes, J. A. (2010): "A concise proof of Kruskal's theorem on tensor decomposition," *Linear Algebra and its Applications*, 432, 1818–1824.

Sasaki, Y. (2015a): "A Contraction Fixed Point Method for Infinite Mixture Models and Direct Counterfactual Analysis," Working Paper, Johns Hopkins University.

Sasaki, Y. (2015b): "Heterogeneity and selection in dynamic panel data," *Journal of Econometrics*, 188, 236 – 249.

Schennach, S. M. (2000): "Estimation of Nonlinear Models with Measurement Error," Working Paper, University of Chicago.

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

Schennach, S. M. (2004b): "Exponential specifications and measurement error," *Economics Letters*, 85, 85–91.

Schennach, S. M. (2004c): "Nonparametric Estimation in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.

Schennach, S. M. (2007a): "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 75, 201–239.

Schennach, S. M. (2007b): "Point Estimation with Exponentially Tilted Empirical Likelihood," *Annals of Statistics*, 35, 634–672.

Schennach, S. M. (2008): "Quantile regression with mismeasured covariates," *Econometric Theory*, 24, 1010–1043.

SCHENNACH, S. M. (2013a): "Measurement Error in Nonlinear Models — A Review," in *Advances in Economics and Econometrics, Econometric society monographs, no. 51*, vol. 3, chap. 8. Cambridge University Press.

SCHENNACH, S. M. (2013b): "Regressions with Berkson errors in covariates — A nonparametric approach," *Annals of Statistics*, 41, 1642–1668.

SCHENNACH, S. M. (2014): "Entropic Latent Variable Integration via Simulation," *Econometrica*, 82, 345–386.

SCHENNACH, S. M. (2016): "Recent Advances in the Measurement Error Literature," *Annual Reviews of Economics*, 8, 341–377.

SCHENNACH, S. M. (2018): "Mismeasured and unobserved variables," in *Handbook of Econometrics*, vol. 7A, invited, under (minor) revision. Elsevier Science.

SCHENNACH, S. M. (2019): "Convolution without independence," *Journal of Econometrics*, 211, 308–318.

SCHENNACH, S. M., AND Y. HU (2013): "Nonparametric Identification and Semiparametric estimation of classical measurement error models without side information," *Journal of the American Statistical Association*, 108, 177–186.

SCHENNACH, S. M., AND V. STARCK (2020): "Using spatial modeling to address covariate measurement error," Working Paper CWP58/20, Cemmap.

SCHENNACH, S. M., H. WHITE, AND K. CHALAK (2012): "Local indirect least squares and average marginal effects in nonseparable structural systems," *Journal of Econometrics*, 166, 282–302.

SCHWARTZ, L. (1966): *Théorie des distributions*. Paris, Hermann.

SHEN, X. (1997): "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555–2591.

SÖDERSTRÖM, T. (2018): *Errors-in-variables methods in system identification*. Springer, Switzerland.

SONG, S., S. SCHENNACH, AND H. WHITE (2015): "Identification and Estimation of Nonseparable Models with Measurement Errors," *Quantitative Economics*, 6, 749–794.

Spearman, C. (1904): "General intelligence, objectively determined and measured," *American Journal of Psychology*, 15, 201–293.

Stoker, T., E. Berndt, D. Ellerman, and S. M. Schennach (2005): "Panel Data Analysis of U.S. Coal Productivity," *Journal of Econometrics*, 127, 131–164.

Stram, D. O., M. Huberman, and A. H. Wu (2002): "Is Residual Confounding a Reasonable Explanation for the Apparent Protective Effects of Beta-carotene Found in Epidemiologic Studies of Lung Cancer in Smokers?," *American Journal of Epidemiology*, 155, 622–628.

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.

Tobin, J. (1969): "A General Equilibrium Approach to Monetary Theory," *J. Money Credit and Banking*, 1, 15–29.

Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50, 945–973.

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.

Wansbeek, T. J., and E. Meijer (2000): *Measurement error and latent variables in econometrics*. Elsevier, New York.

Wilhelm, D. (2015): "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," Working Paper CWP34/15, Cemmap.

Wilhelm, D. (2019): "Testing for the Presence of Measurement Error," Working Paper CWP48/19, Cemmap.

Williams, B. (2019): "Identification of a Nonseparable Model under Endogeneity using Binary Proxies for Unobserved Heterogeneity," *Quantitative Economics*, 10, 527–563.