Identification and Estimation of Linear Dependent Multidimensional Unobservables

Dan Ben-Moshe*
Department of Economics
UCLA

May 18, 2011

Abstract

I consider identification and estimation of nonparametric distributions in a linear panel data model with multidimensional and statistically dependent unobservables. Identification requires no parametric distributional or functional assumptions on each of the unobservables, so they can be unobservable random variables or nonparametric nonseparable functions of covariates and of unobserved heterogeneity. The methods allow for the number of unobservables to be more than twice the number of equations, for some unobservables to be arbitrarily dependent and for variables to be discrete or continuous. The main identifying assumptions are the linearity and panel data structure. The identification techniques can be used to construct a broad class of estimators that are based on empirical first-order partial derivatives of characteristic functions. I show that the estimators are consistent and derive asymptotic convergence rates. I compare the finite sample properties of estimators using Monte Carlo simulations.

*I am extremely grateful to Rosa Matzkin and Jinyong Hahn for the many hours they spent with me from the very beginning of the research process to the writing of this draft. All errors are mine. E-mail: danbm@ucla.edu

1 Introduction

I study identification and estimation of distributions when unobservables are multidimensional and statistically dependent with each unobservable completely nonparametric and nonseparable.[1] Some papers separately deal with multidimensional unobservables, dependence or nonseparability while other papers deal with all of them together but settle for identifying an average effect, imposing some structure on the model or using instrumental variables or control functions. In this paper, linearity is the structural assumption that enables identification. In return, identification is possible even when the number of equations is much less than the number of unobservables and unobservables are highly dependent, nonparametric and nonseparable.

The dimension of unobservables, separability of functions and dependence structure of the random variables crucially influence the degree of difficulty and techniques used for identification and estimation of features in an econometric model. The dimension of unobservables in many applied econometric models is smaller than or equal to the number of equations. The reason is that these models usually require some monotonicity assumption allowing invertibility that may not make sense when there is a relatively large number of unobservables. Maximum likelihood and method of moments estimators, for example, are most easily implemented when functions are invertible so that unobservables can be explicitly expressed in terms of distributional assumptions or moment restrictions.

Separable models of the form $Y = m(X) + \varepsilon$ have the major disadvantage that $\frac{\partial Y}{\partial X}$ fails to capture differences in responses for observationally identical individuals. Some solutions include Heckman et al. (2010) who identify preferences in a nonseparable hedonics model where marginal effects on utility can be heterogeneous. The sacrifice they make for identification is a scalar unobservable heterogeneous term. Chesher (2007) observes that in nonseparable nonparametric models with more unobservables than equations, one must either reduce the dimension of the unobservables, add observable equations or be content to identify average quantities. Chernozhukov et al. (2010) choose to identify average quantities in nonseparable panel data models. The problem is that identifying average quantities may not have useful structural interpretations.

Independence of variables is helpful to identify structural functions. Matzkin (2003) identifies a nonparametric function by observing that the distribution of outcome variables conditional on covariates is the same as the distribution of the unobservable as long as the unobservable and covariates are independent. Independence is helpful but usually considered a strong assumption because choices of covariates and unobservables are often related. When covariates and unobservables are dependent the most common solutions use control functions or instrumental variables. When unobservables are dependent with each other the most common solutions involve

[1] Dependence in this paper is always statistical or linear dependence. It never refers to the dependent variable, which I call the outcome variable.

explicit modeling of the dependence. Altonji and Matzkin (2005) identify average marginal effects by assuming independence of covariates and unobservables conditional on control variables. Torgovitsky (2011) identifies a nonparametric function using instruments and assuming that their effect on the copula is restricted. Cunha et al. (2010) identify a nonparametric production function by explicitly expressing an unobservable in terms of other unobservables and an independent error.

I consider the linear multidimensional panel data model $Y = AU$ where $Y = (Y_1, \ldots, Y_P)'$ is a vector of P measurements, $U = (U_1, \ldots, U_M)'$ is a vector of M unobservables and A is a $P \times M$ matrix of known parameters. This paper is concerned with identification and estimation of the distribution of U. Linearity and panel data are strong assumptions but in return for this structure, identification will be possible even when $P << M$, the unobservables are highly dependent and completely nonseparable. I will present several applications where linearity is a natural or widely accepted assumption.

When $Rank(A) \geq M$ (which indicates that the number of unobservables is smaller than or equal to the number of information-giving equations) and the unobservables are mutually independent then the distribution of U is identified from the distribution of Y and any pseudoinverse, A^+, by $\Pr(U < u) = \Pr(A^+Y < u)$. The linearity and panel data structure in this paper, however, will enable identification even when the number of unobservables is far larger than the number of equations and unobservables are highly dependent. In the examples considered later the number of unobservables will be about twice as large as the number of equations and in one of the examples every unobservable will be arbitrarily dependent with at least one other unobservable.

Notice that the unobservables do not necessarily have to be variables but can be arbitrary nonparametric functions of interest. Once the distribution of the function is identified then the function can be identified by referring to the nonparametric identification literature (see Matzkin (2007) as a reference for many of the standard techniques). Hence, the unobservables U can be unobservable variables as in the deconvolution problem $X = X^* + \varepsilon$ where the true variable X^* is the object of interest or unobservable and nonseparable functions $U_m = g_m(x, \alpha)$ of covariates x and heterogeneity α as in Evdokimov (2011) $Y_{it} = m(X_{it}, \alpha_i) + \varepsilon_{it}$ where the nonparametric function m is identified.

My identification procedure and estimator are closely related the work of Kotlarski (1967), who proved identification in the simplest linear model where unobservables are completely nonparametric and identification is not trivial. The implications of his lemma are ubiquitous in the measurement error literature (see Carroll et al. (2006) and Chen et al. (2011) for detailed reviews on identification and estimation in measurement error models). Ingenious applications of Kotlarski's lemma have also been used by Li et al. (2000) for identification in a common value auction setting and Evdokimov (2011) in a panel data model with nonadditive unobserved heterogeneity and an idiosyncratic shock.

Székely and Rao (2000) generalize Kotlarski's identification lemma to identification in a multi-factor model with independent nonparametric unobservables. They use the $P(P+1)/2$ second-order partial derivatives of the characteristic function to set up a system of equations and prove that a rank condition on a Hessian matrix is necessary and sufficient for nonparametric identification. They can identify up to $M = P(P+1)/2$ unobservables. Bonhomme and Robin (2010) base their estimator on Székely and Rao (2002) system of second-order partial derivatives of characteristic functions and apply their estimator in several economic settings.[2]

Of all these papers, Bonhomme and Robin (2010) is the only one that considers the same general system of equations $Y = AU$ and constructively identifies U. I thus point out some of the important differences between Bonhomme and Robin (2010) and this paper. First, their identification relies crucially on the assumption of mutually independent unobservables.[3] Second, their identification procedure either identifies the entire vector of unobservables or none of them, so there is no possibility for partial identification. Third, their estimator requires stronger moment conditions and nonvanishing characteristic functions. Fourth, their expression of the estimator is substantially different from the one in this paper: they use second-order partial derivatives of characteristic functions while I use first-order partial derivatives of characteristic functions similar to the types of estimators in Kotlarski (1967) and most of the papers on measurement error with repeated measurements. Of the numerous estimators (as a result of overidentification), there are no papers that investigate the issue of a "best" estimator, so the Bonhomme and Robin (2010) estimator may be better for some distributions and worse for others. The theoretical uniform convergence rates suggest that the Bonhomme and Robin (2010) estimator converges slowly when the characteristic function has fat tails. In finite sample simulations their estimator converges fastest in many commonly used distributions (like Standard Normal and Gamma) but is the least robust and does not converge when distributions are multimodal (e.g. Bimodal) or discontinuous (e.g. Uniform). Their estimator takes more CPU time and requires significantly more computer memory due to the need to solve a larger system of equations and hence a larger matrix inversion.

The dimension of unobservables, nonseparability and dependence structure of unobservables have made the analysis of production functions an interesting area of research. Olley and Pakes (1996) and Levinsohn and Petrin (2003) assume the Cobb-Douglas functional form $Y_{it} = \beta_K K_{it} + \beta_L L_{it} + \omega_{it} + \varepsilon_{it}$ where Y_{it} is the log of output, K_{it} is the log of capital input and L_{it} is the log of labor input. ω_{it} represents shocks that are observable or predictable by firms such as managerial ability or defect rates of the manufacturing process while ε_{it} are idiosyncratic shocks. Each equation has two unobservables and is completely separable. ε_{it} is independent of all other variables and ω_{it} and the covariates may be dependent. Because ω_{it} is a scalar, some source of heterogeneity

[2] A replication of their identification procedure is in the appendix.

[3] Bonhomme and Robin (2010) state that it is not straightforward to extend their method to account for dependent unobservables and that it is an interesting problem to research.

(labor-specific effects or capital-specific effects) is ignored. The covariates and ω_{it} are additively separable, so any change in output due to a change in labor or capital fails to capture differences in responses for observationally identical individuals. The dependence between ω_{it} and covariates is solved by modeling the evolution of ω_{it} and the relationship of ω_{it} and covariates. I extend the model to $Y_{ijt} = a_{1t}g(L_{ijt}, \alpha_i) + a_{2t}h(K_{ijt}, \beta_j) + \varepsilon_{ijt}$ where a_{1t} and a_{2t} are unobserved parameters, h and g are unobserved nonparametric functions, α_i and β_j are arbitrarily dependent sources of time-invariant heterogeneity and ε_{ijt} is an idiosyncratic shock.

In addition to the production function application, I will also solve a model of individual earnings dynamics with dependence between unobservables and a measurement error model with repeated measurements focusing on comparing different estimators.

In addition to being concerned with identification, several papers also propose ways to relax the required regularity conditions. The early measurement error literature (and literature on deconvolutions) followed Kotlarski's assumption of nonvanishing characteristic functions. Fan (1991) and Li and Vuong (1998) assumed nonvanishing characteristic functions on finite support while Schennach (2004) assumed nonvanishing characteristic functions on infinite support. Bondesson (1974) was the first to prove identification when the characteristic functions satisfy a "short gap" condition, which meant that the characteristic functions have no intervals of length L for all $L > 0$. More recently, Delaigle et al. (2008), Carrasco and Florens (2010) and Evdokimov and White (2011) impose a similar condition by restricting the characteristic function to have a countable number of isolated zeros. In this paper I impose an absolute continuity condition, which is a weaker condition that includes all the above restrictions as a subset.

The identification strategy consists of three conceptual steps. First, linear combinations of equations are used to simplify the problem in the original space of unobservables $\mathcal{R}^M$. Second, a system of equations is created that is identical to $Y = AU$ except with random variables Y and U replaced by partial derivatives of characteristic functions of random variables. This reduces the problem by creating P systems of equations, each of which lies in the smaller space $\mathcal{R}^P$. Third, a rank condition on a modification of the matrix A ensures that an unobservable partial derivative of a characteristic function can be expressed as a functional of observable partial derivatives of characteristic functions. The characteristic function and density of the unobservable are then recovered by using the fundamental theorem of calculus and a Fourier transform inversion.

The identification results in this paper are constructive, so the estimator replaces population quantities with sample analogs. Unobservables are identified sequentially, so identification and estimation of nuisance random variables can sometimes be avoided. The estimator is easy to implement and requires no numerical optimization. I provide results on the uniform convergence rate of the estimator. The support of distributions can be unbounded and can vanish on sets of Lebesgue measure zero. The convergence rates are relatively slow and depend on the smoothness of observable and unobservable distributions.

The identification strategy produces a broad class of estimators that use partial derivatives of characteristic functions. The choice of estimator in most of the referenced papers is arbitrary. The Monte Carlo Simulations section investigates the finite sample properties of five different estimators in the measurement error model with two measurements. The estimators perform very well in all simulations irrespective of underlying distributions.

The paper is organized as follows. Section 2 presents the model and its assumptions and sets up the empirical illustrations that will later be identified. Section 3 proves identification and provides an algorithm that can be used for identification and construction of estimators. Section 4 constructs the estimator of the characteristic functions and densities. Section 5 proves consistency of the estimator. Section 6 displays Monte Carlo simulations. Section 7 concludes. The Appendix contains some proofs and technical arguments.

2 Model and Assumptions

This section presents and explains the assumptions of the model and three empirical illustrations. The focus of this research is identification and estimation of the distributions of the unobservables as in the following basic setup of a panel data model

Assumption 1. *$\{Y_n, U_n\}_{n=1}^N$ is a random independent sample from the random vector (Y, U), where $Y_n = (Y_{n1}, \ldots Y_{nP})$ is observed and $U_n = (U_{n1}, \ldots U_{nM})$ is unobserved. Let a_{pm}, $p = 1, \ldots, P$, $m = 1, \ldots, M$ be real valued scalars that are known to the researcher.[4] Assume*

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_P \end{pmatrix} = \begin{pmatrix} a_{11}U_1 + \ldots + a_{1M}U_M \\ \vdots \quad \ddots \quad \vdots \\ a_{P1}U_1 + \ldots + a_{PM}U_M \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & \ldots & a_{1M} \\ \vdots & \ddots & \ddots & \vdots \\ a_{11} & \ldots & \ldots & a_{PM} \end{pmatrix} \begin{pmatrix} U_1 \\ \vdots \\ U_M \end{pmatrix} \tag{1}$$

This is compactly written as $Y = AU$.

As is conventional in economic models, Y denotes the outcome variables and U the unobservable variables but unlike most economic models there seem to be no covariates, X. This is justified in a few ways. First, covariates can enter the model in a known way and included as part of Y. Second, in some applications identification of the distribution of an unobservable is a first step within a larger problem. For example, Cunha, et al. (2010) consider a multistage model where they identify the distribution of latent variables like parental investment in the first stage and then use the identified distributions along with other covariates to identify a production function in the second stage. Third, in some applications, conditional on covariates, the model becomes a

[4]In applications where p represents time, $p = 1, \ldots, P$ is usually denoted by $t = 1, \ldots, T$.

system of equations where the unobservables are nonparametric and nonseparable functions of covariates and an unobservable random scalar. For example, Evdokimov (2011) considers the model $Y_{it} = m(X_{it}, \alpha) + \varepsilon_{it}$. Conditional on $X_{it} = x$, let $U_1 = m(x, \alpha)$ and $U_2 = \varepsilon_{it}$ then the model is written as $Y = U_1 + U_2$.

Identification relies crucially on linearity and a known or consistent estimator of the matrix A.[5][6] These are strong, but in some applications, natural or widely used assumptions. For example, in the earnings dynamics model considered later, observed total income is equal to the sum of observable and unobservable income components. Thus, linearity and the known matrix A (consisting of 1s and 0s) follow by definition. In the measurement error literature the most common assumption used is that the observed mismeasured variable equals the sum of the unobserved true variable and an unobserved measurement error. The idea originates from the signal processing literature where an observed signal (the mismeasured variable) is an unknown signal (true variable) corrupted by additive noise. Recently, Hu and Schennach (2007) weaken the linearity assumption in a measurement error context. In the empirical illustrations below, the form of A is natural, widely used or identified as a preliminary step.

The dependence structure of the unobservable U is as follows:

Assumption 2. *Assume* $U = (U_1, \ldots, U_M) = \left(Z_1, \ldots, Z_{M_{ind}}, W_1, \ldots, W_{M_{dep}}\right)$, $M = M_{ind} + M_{dep}$ *where* U_m, $m = 1, \ldots, M$, *are scalar real-valued nondegenerate random variables,* $Z = (Z_1, \ldots, Z_{M_{ind}})$ *is a vector of mutually independent random variables,* $W = \left(W_1, \ldots, W_{M_{dep}}\right)$ *is a vector of dependent random variables and* Z *is independent of* W.

The model can be generalized to $(U_1, \ldots, U_M) = \left(Z_1, \ldots, Z_{M_{ind}}, W_1^1, \ldots, W_{M_1}^1, \ldots, W_1^{\widetilde{M}_{dep}}, \ldots, W_{M_{\widetilde{M}_{dep}}}^{\widetilde{M}_{dep}}\right)$, $M = M_{ind} + \sum_{j=1}^{\widetilde{M}_{dep}} M_j$ where $Z = (Z_1, \ldots, Z_{M_{ind}})$ is a vector of mutually independent random variables and $W^j = \left(W_1^j, \ldots, W_{M_j}^j\right)$ is a vector of dependent random variables. $Z, W^1, \ldots, W^{\widetilde{M}_{dep}}$ are mutually independent. All the results continue to hold in this more general setting but to avoid unnecessarily complicating notation, identification is proved under Assumption 2. The empirical illustration described later in this section on earnings dynamics considers $U = (W^1, W^2)$.

Definition 1. *Denote* $Q_m^{U,p^*,m^*} = \{\{a_{p^*m \neq 0}\} \cup \{U_m \text{ and } U_{m^*} \text{ are dependent }\}\}$. *For any matrix* $A = (A_1 \ \ldots \ A_M)$ *with columns* A_m *and entries* a_{pm}, $p = 1, \ldots, P$, $m = 1, \ldots, M$ *define*

$$A^{U,p^*,m^*} = \left(A_1 \mathbf{I}\left(Q_1^{U,p^*,m^*}\right) \ \ldots \ A_M \mathbf{I}\left(Q_M^{U,p^*,m^*}\right)\right)' = \begin{pmatrix} a_{11}\mathbf{I}\left(Q_1^{U,p^*,m^*}\right) & \ldots & a_{P1}\mathbf{I}\left(Q_1^{U,p^*,m^*}\right) \\ \vdots & \ddots & \vdots \\ a_{1M}\mathbf{I}\left(Q_M^{U,p^*,m^*}\right) & \ldots & a_{PM}\mathbf{I}\left(Q_M^{U,p^*,m^*}\right) \end{pmatrix}$$

[5] In future work I hope to generalize the model to $Y = g(U)$ where g is a known multivariate function.

[6] The problem looks like one from regression analysis except the objective is to identify the distribution of U instead of the parameters A.

where I_E is the indicator function.[7]

An unobservable U_m (a column in A and a row in A^{U,p^*,m^*}) is included in A^{U,p^*,m^*} if its coefficient in row p^* is nonzero (in the Identification section, identification will come from "moving" equation p^* so I will refer to equation p^* as the source from where identification is coming) or U_m and U_{m^*} are dependent. Hence, A^{U,p^*,m^*} is the transpose of the matrix A and only includes unobservables that are in some way related to the identification of U_{m^*}. When $U_m \in Z$ then U_m and U_{m^*} are independent so $Q_m^{U,p^*,m^*} = \{a_{p^*m \neq 0}\}$ which means that U_m can only affect identification through equation p^*. When $U_m \in W$ then U_m and U_{m^*} are dependent if and only if $U_{m^*} \in W$ so $Q_m^{U,p^*,m^*} = \{\{a_{p^*m \neq 0}\} \cup \{U_{m^*} \in W\}\}$ which means that U_m can affect identification through equation p^* or through its dependence with U_{m^*}. Hence, the superscript U is the vector of random variables for the dependence structure, the superscript p^* indicates that equation p^* will be the source of identification and the superscript m^* indicates that the distribution of U_{m^*} will be identified. Construction of A^{U,p^*,m^*} will be illustrated later in this section.

To identify the distribution of U_{m^*}, the assumptions on the matrix A and the assumptions on the moments of U are as follows:

Assumption 3. *There exists a matrix B and $p^* \in \{1, \ldots, P\}$ such that*

i. $\widetilde{Y} = \widetilde{A}U$ *where* $\widetilde{Y} = BY$ *and* $\widetilde{A} = BA$.

ii. $e^{m^*} = \left(e_1^{m^*}, \ldots, e_P^{m^*}\right)'$ *solves* $\widetilde{A}^{U,p^*,m^*} e^{m^*} = e_{m^*}$.[8]

iii. Denote $W_{-p^*-m} = \{W_{m'} | \ m' \neq m, \ \widetilde{a}_{p^*m} \neq 0 \text{ and } \widetilde{a}_{p^*m'} = 0\}$. *Assume* $E[W_m|W_{-p^*-m}] = E[W_m]$, $E[W_m|U_{m^*}] = E[W_m]$ *and* $E[Z_m]$ *are known. Assume* $E[|U_m|]$ *is finite.*

iv. Let $\phi_{Y'e^{m^*}}(s)$ *be the characteristic function of* $Y'e^{m^*} = \sum_{p=1}^P e_p^{m^*} Y_p$ *evaluated at s and* $\phi_{m^*}(s)$ *be the characteristic function of* U_{m^*} *evaluated at s.*[9] *Define the measures* $\mu = \int |\phi_{Y'e^{m^*}}(s)| \mathrm{d}s$ *and* $\nu = \int |\phi_{m^*}(s)| \mathrm{d}s$. *Assume that for every set X with nonzero Lebesgue measure* $\mu(X) = 0$ *implies* $\nu(X) = 0$.[10] *For all intervals* $(a, b) \subset (-\infty, \infty)$ *that do not contain any such X assume* $\int_a^b \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u < \infty$.

A is not invertible, but left multiplication by B as in Assumption 3i. gets as close as possible to an inversion by transforming A to reduced row echelon form with U_{m^*} as a free variable. This results in U_{m^*} in as many equations as possible and a sparse matrix.

[7] The indicator function, $\mathbf{I}(E)$, is defined as

$$\mathbf{I}(E) = \begin{cases} 1 & \text{If E is true} \\ 0 & \text{Otherwise} \end{cases}$$

[8] $e_{m^*} = (0, \ldots, 0, 1, 0, \ldots, 0)'$ where 1 is in the m^{*th} coordinate.

[9] The characteristic function of a random vector V is $\phi_V(t) = E[\exp(iV't)]$.

[10] ν is absolutely continuous with respect to μ, which is written as $\nu << \mu$.

The span of A is smaller than the dimension of unobservables ($\mathcal{R}^P \subset \mathcal{R}^M$), so after left multiplication by B in Assumption 3i., I will reduce the problem from $\mathcal{R}^M$ to $\mathcal{R}^P$ (by a Fourier transformation and differentiation). In the smaller space $\mathcal{R}^P$, $\widetilde{A}^{U,p^*,m^*}$ as in Assumption 3ii. will be the object used to take linear combinations of the P observables, $Y_1, \dots Y_P$ and span $\mathcal{R}^P$. The rank condition ensures that U_{m^*} lies in this space.

Both left multiplication of A by B and right multiplication of A^{U,p^*,m^*} by e^{m^*} take linear combinations of rows of A. I emphasize again the important difference: A includes all unobservables while $A^{p^*m^*}$, whose coefficients are multiplied by an indicator variable, only includes unobservables that have a connection to the source of identification (which is equation p^*) through $\widetilde{a}_{p^*m} \neq 0$ or the object to be identified (which is U_{m^*}) through $U_m \cap U_{m^*} \in W$.

Assumption 3iii. is a location normalization and assumes that some first moments are finite. Many economic models assume an error term with mean (or conditional mean) equal to zero, which is enough for Assumption 3iii. to be satisfied.

To understand Assumption 3iv., I need to explain the connection between characteristic functions and densities and between the unobservable characteristic function and observable characteristic function. The density of a random variable U_{m^*} at u is uniquely expressed as

$$f_{m^*}(u) = \frac{1}{2\pi} \int e^{-isu} \phi_{m^*}(s) \mathrm{d}s$$

where $\left\{e^{-isu},\ s \in \mathcal{R}\right\}$ is an infinite collection of basis functions in $L^2(\mathcal{R})$ and the characteristic functions, $\phi_{m^*}(s)$, are weights at each of the basis functions. The unsigned measure, $\nu = \int |\phi_{m^*}(s)| \mathrm{d}s$, assigns lengths to sets. If $\nu(X)$ equals zero then no weight is given to basis functions in X (i.e. $\int_X e^{-isu} \phi_{m^*}(s) \mathrm{d}s = 0$). Similarly, the density of a random variable $Y' e^{m^*}$ is uniquely expressed in terms of basis functions $\left\{e^{-isu},\ s \in \mathcal{R}\right\}$ and weights $\phi_{Y' e^{m^*}}(s)$ and a measure is defined as $\mu = \int |\phi_{Y' e^{m^*}}(s)| \mathrm{d}s$.

In the Identification section, $\phi_{m^*}(s)$ will be expressed as a functional of $\phi_{Y' e^{m^*}}$ that will be undefined when $\phi_{Y' e^{m^*}}$ equals zero on sets of nonzero Lebesgue measure. A natural solution is to require $\phi_{m^*} = 0$ on these sets; this is exactly the condition $\mu(X) = 0$ implies $\nu(X) = 0$. Intuitively, the density we need to know the value of ϕ_{m^*} on $\mathcal{R}$, but on sets where $\phi_Y = 0$ we do not know ϕ_{m^*}, so we assume that $\phi_{m^*} = 0$ on these sets.

Assumption 3iv., includes as special cases characteristic functions that do not vanish on their support (bounded or unbounded), characteristic functions with isolated zeros and characteristic functions that have gaps.

To facilitate exposition, the panel data models of three empirical illustrations are presented. The purpose of these examples is to demonstrate different types of problems, dependence assumptions and identifying techniques through various choices of B, p^* and e^{m^*}. I will show that each example conforms to the setup in Assumption

1, introduce the dependence structure of the unobservables (Assumption 2,) and show that Assumption 3 is satisfied. The emphasis will be on constructing $\widetilde{A}^{U,p^*,m^*}$ and showing that Assumption 3ii. holds. In the Identification section, I return to the empirical illustrations and prove identification.

Example 1: Measurement Error With Two Measurements

Consider the measurement error model with repeated measurements as in Li and Vuong (1998)

$$X_1 = X^* + \varepsilon_1$$
$$X_2 = X^* + \varepsilon_2$$

where X_1 and X_2 are observed measurements, X^* is an unobserved true variable and ε_1 and ε_2 are measurement errors. Let $Y = (X_1, X_2)'$ and $U = (X^*, \varepsilon_1, \varepsilon_2)'$ then

$$Y = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} U$$

Assume $E[X^*|\cdot] = E[\varepsilon_1|\cdot] = E[\varepsilon_2|\cdot] = 0$ where "$|\cdot$" denotes "conditional on all other random variables" (this notation is also used in the following examples) and X^*, ε_1 and ε_2 are mutually independent.[11]

Only the distribution of X^* is interesting. I verify Assumptions 3i.-iii. for three different choices of B, p^* and m^* that are used in three different estimators. Assumption iii. always holds because $E[X^*|\cdot] = E[\varepsilon_1|\cdot] = E[\varepsilon_2|\cdot] = 0$.

Estimator A. Let $B = I_2$ so $\widetilde{A} = A$.[12] $U_1 = X^*$ is identified from the first equation . Hence, the source of identification is equation 1 ($p^* = 1$) and the unobservable to be identified is U_1 ($m^* = 1$). I now construct $\widetilde{A}^{U,1,1}$. $U_1 = X^*$ and $U_2 = \varepsilon_1$ play a part in the identification process because they have nonzero coefficients in equation $p^* = 1$ so they are included in $\widetilde{A}^{U,1,1}$. ε_2 on the other hand is not in equation $p^* = 1$ nor are X^* and ε_2 dependent so ε_2 (the last column of A) is not included in $\widetilde{A}^{U,1,1}$. To the right of $\widetilde{A}^{U,1,1}$ I label the variable and include the reason the reason why it is included in $\widetilde{A}^{U,1,1}$. When $(p^*, m^*) = (1, 1)$ then

$$\widetilde{A}^{U,1,1} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{array}{ll} \leftarrow & X^* \text{ is the random variable to be identified} \\ \leftarrow & \varepsilon_1 \text{ is in equation } p^* = 1 \\ \leftarrow & \varepsilon_2 \text{ plays no part in the identification process} \end{array}$$

[11] These conditions are stronger than what is needed for identification. In fact, U is identified under the weaker assumptions $E[\varepsilon_1|X^*] = 0$, $E[\varepsilon_2] = 0$ and (X^*, ε_1) independent of ε_2. The goal of this example, however, is not to find the weakest set of conditions for identification but rather to compare different estimators (only one estimator is consistent with the dependence assumption). One of the simulations in the Monte Carlo Simulations section assumes the weaker dependence assumption.

[12] I_K is the identity matrix of size K.

Finally, when $e^1 = (0,1)'$ then $\widetilde{A}^{U,1,1}e^1 = e_1$ so Assumption 3ii. is satisfied. Hence, Assumptions 3i.-iii. are satisfied for identification of $U_1 = X^*$.

Estimator B. The distribution of ε_1 is identified as a preliminary step to identifying the distribution of X^*. Let $B = I_2$ so $\widetilde{A} = A$. $U_2 = \varepsilon_1$ ($m^* = 2$) will be identified from the first equation ($p^* = 1$). When $(p^*, m^*) = (1,2)$ and $e^2 = (1,-1)$ then

$$\widetilde{A}^{U,1,2} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{array}{l} \leftarrow \quad X^* \text{ is in equation } p^* = 1 \\ \leftarrow \quad \varepsilon_1 \text{ is the random variable to be identified} \\ \leftarrow \quad \varepsilon_2 \text{ plays no part in the identification process} \end{array} \qquad \widetilde{A}^{U,1,2}e^2 = e_2$$

Assumptions 3i.-iii. are satisfied for identification of $U_2 = \varepsilon_1$. X^* will be identified by the deconvolution of $Y_1 = X^* + \varepsilon_1$ which is possible because ε_1 is independent of X^*, the distribution of Y_1 is observed and the distribution of ε_1 is identified.

Estimator C. As in Estimator B, the distribution of ε_1 is identified as a preliminary step to identifying the distribution of X^*. Let

$$B = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \qquad \widetilde{A} = BA = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

$\varepsilon_1 = U_2$ ($m^* = 2$) is identified from the second equation ($p^* = 2$). When $(p^*, m^*) = (2,2)$ and $e^2 = (1,0)$ then

$$\widetilde{A}^{U,2,2} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 0 & -1 \end{pmatrix} \begin{array}{l} \leftarrow \quad X^* \text{ plays no part in the identification process} \\ \leftarrow \quad \varepsilon_1 \text{ is the random variable to be identified} \\ \leftarrow \quad \varepsilon_2 \text{ is in equation } p^* = 2 \end{array} \qquad \widetilde{A}^{U,2,2}e^2 = e_2$$

Assumptions 3i.-iii. are satisfied for identification of $U_2 = \varepsilon_1$. As in Estimator B, identification of X^* follows by deconvolution.

Example 2: Earnings Dynamics

Consider the earnings dynamics model as in Bonhomme and Robin (2010) with weaker dependence assumptions

$$\begin{aligned} w_{nt} &= f_n + y^P_{nt} + y^T_{nt}, && n = 1, \ldots, N,\ t = 1, \ldots, T \\ y^P_{nt} &= y^P_{nt-1} + \varepsilon_{nt} && t \geq 2 \\ y^T_{nt} &= \eta_{nt} \end{aligned}$$

$$\eta_{n1} = \eta_{nT} = 0$$

where w_{nt} is the residual of a regression of individual log earnings, f_n is a fixed effect, y^T_{nt} is transitory earnings, y^P_{nt} is persistent earnings, ε_{nt} is a persistent earnings shock and and η_{nt} is a transitory earnings shock. Set $T = 4$ and difference out the individual fixed effects f_n by $w_{nt} - w_{nt-1}$. Let $Y = (w_{n2} - w_{n1}, w_{n3} - w_{n2}, w_{n4} - w_{n3})'$ and $U = (\eta_{n2}, \eta_{n3}, \varepsilon_{n2}, \varepsilon_{n3}, \varepsilon_{n4})'$ then

$$Y = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} U$$

Assume $E[\eta_{n2}|\cdot] = E[\eta_{n3}] = E[\varepsilon_{n2}|\cdot] = E[\varepsilon_{n3}|\cdot] = E[\varepsilon_{n4}|\cdot] = 0$ and (η_{n2}, η_{n3}) and $(\varepsilon_{n2}, \varepsilon_{n3}, \varepsilon_{n4})$ are independent.[13] I verify Assumptions 3i.-iii. Assumption iii. always holds because $E[\eta_{n2}|\cdot] = E[\eta_{n3}] = E[\varepsilon_{n2}|\cdot] = E[\varepsilon_{n3}|\cdot] = E[\varepsilon_{n4}|\cdot] = 0$.

Let $B = I_5$ so $\widetilde{A} = A$. $U_1 = \eta_{n2}$ $(m^* = 1)$ is identified from the second equation $(p^* = 2)$. When $(p^*, m^*) = (2, 1)$ and $e^1 = (1, 0, 0)$ then

$$\widetilde{A}^{U,2,1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{array}{l} \leftarrow \quad \eta_{n2} \text{ is the random variable to be identified} \\ \leftarrow \quad \eta_{n3} \text{ is in equation } p^* = 2 \\ \leftarrow \quad \varepsilon_{n2} \text{ plays no part in the identification process} \\ \leftarrow \quad \varepsilon_{n3} \text{ is in equation } p^* = 2 \\ \leftarrow \quad \varepsilon_{n4} \text{ plays no part in the identification process} \end{array} \qquad \widetilde{A}^{U,2,1} e^1 = e_1$$

Hence, Assumption 3ii. is satisfied for identification of η_{n2}. $U_2 = \eta_{n3}$ $(m^* = 2)$ and $U_4 = \varepsilon_{n3}$ $(m^* = 4)$ are also identified from equation $p^* = 2$ so $\widetilde{A}^{U,2,2} = \widetilde{A}^{U,2,4} = \widetilde{A}^{U,2,1}$. When $e^2 = (0, 0, -1)$ then $\widetilde{A}^{U,2,2} e^2 = e_2$ and when $e^4 = (1, 1, 1)$ then $\widetilde{A}^{U,2,4} e^4 = e_4$ so Assumption 3ii. is also satisfied for η_{n3} and ε_{n3}. Hence, Assumptions 3i.-iii. are satisfied for identification of $U_1 = \eta_{n2}$, $U_2 = \eta_{n3}$ and $U_4 = \varepsilon_{n3}$.

Now let

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad \widetilde{A} = BA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

[13] The unobservables are identified under other dependency assumptions.

$U_3 = \varepsilon_{n2}$ ($m^* = 3$) is identified from the third equation ($p^* = 3$). When $(p^*, m^*) = (3,3)$, $e^3 = (1,0,0)$ then

$$\widetilde{A}^{U,3,3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{array}{l} \leftarrow \eta_{n2} \text{ plays no part in the identification process} \\ \leftarrow \eta_{n3} \text{ plays no part in the identification process} \\ \leftarrow \varepsilon_{n2} \text{ is in equation } p^* = 2 \\ \leftarrow \varepsilon_{n3} \text{ is the random variable to be identified} \\ \leftarrow \varepsilon_{n4} \text{ is in equation } p^* = 2 \end{array} \qquad \widetilde{A}^{U,3,3} e^3 = e_3$$

Hence, Assumption 3ii. is satisfied for identification of ε_{n2}. $U_5 = \varepsilon_{n4}$ ($m^* = 5$) is also identified from equation $p^* = 3$ so $\widetilde{A}^{U,3,5} = \widetilde{A}^{U,3,3}$. When $e^5 = (0,1,0)$ then $\widetilde{A}^{U,3,5} e^5 = e_5$ so Assumption 3ii. is also satisfied for ε_{n4}. Hence, Assumptions 3i.-iii. are satisfied for identification of $U_3 = \varepsilon_{n2}$ and $U_5 = \varepsilon_{n4}$.

Example 3: Production Function

Consider a Cobb-Douglas-type production function

$$Y_{ijt} = a_t g(L_{ijt}, \alpha_i) + b_t h(K_{ijt}, \beta_j) + \varepsilon_{ijt}$$

where Y_{ijt} is the log of output, L_{ijt} is the log of labor, K_{ijt} is the log of capital, α_i is unobserved labor heterogeneity and β_j is unobserved capital heterogeneity. Assume a and b are known and $a \neq b$.[14] Suppose the number of labor-capital, (i,j), observations is large and $T = 2$. Let $L_{ij1} = L_{ij'2} = l$ and observe that for the same labor-individual $g(L_{ij1}, \alpha_i) = g(L_{ij'2}, \alpha_i) = g(l, \alpha_i)$. Similarly let $L_{i'j'1} = L_{i'j2} = \bar{l}$ then $g(L_{i'j'1}, \alpha_{i'}) = g(L_{i'j2}, \alpha_{i'}) = g(\bar{l}, \alpha_{i'})$. Let $K_{ij1} = K_{i'j2} = k$ and observe that for the same capital-individual $h(K_{ij1}, \beta_j) = h(K_{i'j2}, \beta_j) = h(k, \beta_j)$. Similarly let $K_{i'j'1} = K_{ij'2} = \bar{k}$ then $h(K_{i'j'1}, \beta_{j'}) = h(K_{ij'2}, \beta_{j'}) = h(\bar{k}, \beta_{j'})$.[15]

$$\begin{aligned} Y_{ij1} &= g(l, \alpha_i) + h(k, \beta_j) + \varepsilon_{ij1} \\ Y_{ij'2} &= ag(l, \alpha_i) + bh(\bar{k}, \beta_{j'}) + \varepsilon_{ij'2} \\ Y_{i'j'1} &= g(\bar{l}, \alpha_{i'}) + h(\bar{k}, \beta_{j'}) + \varepsilon_{i'j'1} \\ Y_{i'j2} &= ag(\bar{l}, \alpha_{i'}) + bh(k, \beta_j) + \varepsilon_{i'j2} \end{aligned}$$

[14] They are identified in the Identification section.

[15] This happens when there are two labor-individuals and two capital-individuals. At $T = 1$ they are randomly grouped together and at $T = 2$ the groupings are switched.

Let $Y = (Y_{ij1}, Y_{ij'22}, Y_{i'j'1}, Y_{i'j2})'$ and $U = \left(g(l,\alpha_i), g(\bar{l},\alpha_{i'}), h(k,\beta_j), h(\bar{k},\beta_{j'}), \varepsilon_{ij1}, \varepsilon_{ij'2}, \varepsilon_{i'j'1}, \varepsilon_{i'j2}\right)'$ then

$$Y = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ a & 0 & 0 & b & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & a & b & 0 & 0 & 0 & 0 & 1 \end{pmatrix} U$$

Assume $E[g(l,\alpha_i)|\cdot] = E[g(\bar{l},\alpha_{i'})|\cdot] = E[h(k,\beta_j)|\cdot] = E[h(\bar{k},\beta_{j'})|\cdot] = E[\varepsilon_{ij1}|\cdot] = E[\varepsilon_{ij'2}|\cdot] = E[\varepsilon_{i'j'1}|\cdot] = E[\varepsilon_{i'j2}|\cdot] = 0$ and ε_{ij1}, $\varepsilon_{ij'2}$, $\varepsilon_{i'j'1}$, $\varepsilon_{i'j2}$ and $\left(g(l,\alpha_i), g(\bar{l},\alpha_{i'}), h(k,\beta_j), h(\bar{k},\beta_{j'})\right)$ are mutually independent.[16] I verify Assumptions 3i.-iii. Assumption iii. always holds because $E[g(l,\alpha_i)|\cdot] = E[g(\bar{l},\alpha_{i'})|\cdot] = E[h(k,\beta_j)|\cdot] = E[h(\bar{k},\beta_{j'})|\cdot] = E[\varepsilon_{ij1}|\cdot] = E[\varepsilon_{ij'2}|\cdot] = E[\varepsilon_{i'j'1}|\cdot] = E[\varepsilon_{i'j2}|\cdot] = 0$.

Let

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -b & \frac{b}{a} & -\frac{b^2}{a} & 1 \end{pmatrix} \qquad \widetilde{A} = BA = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ a & 0 & 0 & b & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & a - \frac{b^2}{a} & 0 & 0 & -b & \frac{b}{a} & -\frac{b^2}{a} & 1 \end{pmatrix}$$

$U_5 = \varepsilon_{ij1}$ ($m^* = 5$) is identified from the fourth equation ($p^* = 4$). When $(p^*, m^*) = (4,5)$ and $e^5 = (1,0,0,0)$ then

$$\widetilde{A}^{U,4,5} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{a^2-b^2}{a} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -b \\ 0 & 1 & 0 & \frac{b}{a} \\ 0 & 0 & 1 & -\frac{b^2}{a} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{l} \leftarrow \quad g(l,\alpha_i) \text{ plays no part in the identification process} \\ \leftarrow \quad g(\bar{l},\alpha_{i'}) \text{ is in equation } p^* = 4 \\ \leftarrow \quad h(k,\beta_j) \text{ plays no part in the identification process} \\ \leftarrow \quad h(\bar{k},\beta_{j'}) \text{ plays no part in the identification process} \\ \leftarrow \quad \varepsilon_{ij1} \text{ is the random variable to be identified} \\ \leftarrow \quad \varepsilon_{ij'2} \text{ is in equation } p^* = 4 \\ \leftarrow \quad \varepsilon_{i'j'1} \text{ is in equation } p^* = 4 \\ \leftarrow \quad \varepsilon_{i'j2} \text{ is in equation } p^* = 4 \end{array} \qquad \widetilde{A}^{U,4,5} e^5 = e_5$$

Assumptions 3i.-iii. are satisfied for identification of $U_5 = \varepsilon_{ij1}$. $U_6 = \varepsilon_{ij'2}$ ($m^* = 6$) is also identified from equation $p^* = 4$ so $\widetilde{A}^{U,4,5} = \widetilde{A}^{U,4,6}$. When $e^6 = (0,1,0,0)$ then $\widetilde{A}^{U,4,6} e^6 = e_6$ so Assumption 3ii. is also satisfied for $\varepsilon_{ij'2}$. Hence, Assumptions 3i.-iii. are satisfied for identification of $U_5 = \varepsilon_{ij1}$ and $U_6 = \varepsilon_{ij'2}$.

[16]The unobservables are identified under other dependency assumptions.

Let

$$B=\begin{pmatrix}1 & -\frac{a}{b^2} & \frac{a}{b} & -\frac{1}{b}\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \qquad \widetilde{A}=BA=\begin{pmatrix}1-\frac{a^2}{b^2} & 0 & 0 & 0 & 1 & -\frac{a}{b^2} & \frac{a}{b} & -\frac{1}{b}\\ a & 0 & 0 & b & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0\\ 0 & a & b & 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

$U_7=\varepsilon_{i'j'2}$ $(m^*=7)$ is identified from the first equation $(p^*=1)$. When $(p^*,m^*)=(1,7)$ and $e^7=(0,0,1,0)$ then

$$\widetilde{A}^{U,1,7}=\begin{array}{c}\begin{pmatrix}1-\frac{a^2}{b^2} & a & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ -\frac{a}{b^2} & 1 & 0 & 0\\ \frac{a}{b} & 0 & 1 & 0\\ -\frac{1}{b} & 0 & 0 & 1\end{pmatrix}\end{array}\begin{array}{l}\leftarrow\quad g(l,\alpha_i)\ \text{is in equation } p^*=1\\ \leftarrow\quad g(\bar{l},\alpha_{i'})\ \text{plays no part in the identification process}\\ \leftarrow\quad h(k,\beta_j)\ \text{plays no part in the identification process}\\ \leftarrow\quad h(\bar{k},\beta_{j'})\ \text{plays no part in the identification process}\\ \leftarrow\quad \varepsilon_{ij1}\ \text{is in equation } p^*=1\\ \leftarrow\quad \varepsilon_{ij'2}\ \text{is in equation } p^*=1\\ \leftarrow\quad \varepsilon_{i'j'1}\ \text{is the random variable to be identified}\\ \leftarrow\quad \varepsilon_{i'j2}\ \text{is in equation } p^*=1\end{array} \qquad \widetilde{A}^{U,1,7}e^7=e_7$$

Assumptions 3i.-iii. are satisfied for identification of $U_7=\varepsilon_{i'j'1}$. $U_8=\varepsilon_{i'j2}$ $(m^*=8)$ is also identified from equation $p^*=1$ so $\widetilde{A}^{U,1,8}=\widetilde{A}^{U,1,7}$. When $e^8=(0,0,0,1)$ then $\widetilde{A}^{U,1,8}e^8=e_8$ so Assumption 3ii. is also satisfied for $\varepsilon_{i'j2}$. Hence, Assumptions 3i.-iii. are satisfied for identification of $U_7=\varepsilon_{i'j'1}$ and $U_8=\varepsilon_{i'j2}$

Because ε_{ij1}, $\varepsilon_{ij'2}$, $\varepsilon_{i'j'2}$ and $\varepsilon_{i'j2}$ are identified and independent of $(g(l,\alpha_i),g(\bar{l},\alpha_{i'}),h(k,\beta),h(\bar{k},\beta_{j'}))$, a new system of equations is setup and Assumption 3i-ii. are checked,

$$\begin{pmatrix}Y_1-U_5\\ Y_2-U_6\\ Y_4-U_7\\ Y_4-U_8\end{pmatrix}=\begin{pmatrix}1 & 0 & 1 & 0\\ a & 0 & 0 & b\\ 0 & 1 & 0 & 1\\ 0 & a & b & 0\end{pmatrix}\begin{pmatrix}U_1\\ U_2\\ U_3\\ U_4\end{pmatrix}$$

Let

$$B = \frac{1}{a^2 - b^2} \begin{pmatrix} -b^2 & a & -ab & b \\ -ab & b & -b^2 & a \\ a^2 & -a & ab & -b \\ ab & -b & a^2 & -a \end{pmatrix} \qquad \widetilde{A} = BA = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$U_1 = g(l, \alpha)$ $(m^* = 1)$ is identified from the first $(p^* = 1)$ equation. When $(p^*, m^*) = (1,1)$ and $e^1 = (1,0,0,0)$

$$\widetilde{A}^{U,1,1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{array}{l} \leftarrow \quad g(l, \alpha_i) \text{ is the random variable to be identified} \\ \leftarrow \quad g(\bar{l}, \alpha_{i'}) \text{ is dependent with } g(l, \alpha_i) \\ \leftarrow \quad h(k, \beta_j) \text{ is dependent with } g(l, \alpha_i) \\ \leftarrow \quad h(\bar{k}, \beta_{j'}) \text{ is dependent with } g(l, \alpha_i) \end{array} \qquad \widetilde{A}^{U,1,1} e^1 = e_1$$

Assumptions 3i.-iii. are satisfied for identification of $U_1 = g(l, \alpha_i)$. Similarly, $U_2 = g(\bar{l}, \alpha_{i'})$, $U_3 = h(k, \beta_j)$ and $U_4 = h(\bar{k}, \beta_{j'})$ are identified from equations $p^* = 2$, $p^* = 3$ and $p^* = 4$ respectively so $\widetilde{A}^{U,2,2} = \widetilde{A}^{U,2,4} = \widetilde{A}^{U,2,1} = \widetilde{A}^{U,1,1}$. When $e^2 = (0,1,0,0)$ then $\widetilde{A}^{U,2,2} e^2 = e_2$, when $e^3 = (0,0,1,0)$ then $\widetilde{A}^{U,3,3} e^3 = e_3$ and when $e^4 = (0,0,0,1)$ then $\widetilde{A}^{U,4,4} e^4 = e_4$ so Assumption 3ii. is also satisfied for $g(\bar{l}, \alpha_{i'})$, $h(k, \beta_j)$ and $h(\bar{k}, \beta_{j'})$. Hence, Assumptions 3i.-iii. are satisfied for identification of $U_1 = g(l, \alpha_i)$, $U_2 = g(\bar{l}, \alpha_{i'})$, $U_3 = h(k, \beta_j)$ and $U_4 = h(\bar{k}, \beta_{j'})$.

Unobservables are identified separately but, as in some of the examples above, $\widetilde{A}^{U,p^*,m_1^*} = \widetilde{A}^{U,p^*,m_2^*}$ so that the same matrix is used to identify the distributions of $U_{m_1^*}$ and $U_{m_2^*}$. The sequential identification technique implies that sometimes U is partially identified and sometimes nuisance unobservables are not identified at all (like in Example 1 where identification of ε_2 was avoided).

Additionally, some unobservables are identified using Assumption 3 but others fail Assumption 3 (like in Example 1 Estimator B, Example 1 Estimator C and Example 3); then a smaller system of equations is created and identification may be possible from this new system. A recursive (and terminating) technique can be used to identify unobservables but possibly not all of them.

I briefly mention a few other applications of the model as set up Assumption 1. The measurement error model with repeated measurements can be generalized to

$$X_p = \sum_{m=1}^{M} X_m^* \mathbf{I}(S_{pm}) + \varepsilon_p \qquad p = 1, \ldots, P$$

where X_p, $p = 1, \ldots, P$ are P observed measurements, X_m^*, $m = 1, \ldots, M$ are M unobserved true variables,

$\mathbf{I}(S_{pm})$ is an indicator that X_m^* is included in equation p and ε_p, $p = 1, \ldots, P$ are measurement errors. Different bureaus collect information on different segments of population income, census counts etc. The information can be combined using the techniques in this paper.

Li et al. (2000) use the results from the measurement error literature and a solution mechanism for a first price auction to identify distributions when each bidder has valuation $U_0 + A_j$, $j = 1, \ldots, L$ where U_0 is the "common" value, A_j is an idiosyncratic shock and L is the number of bidders. This can be extended in a similar way to the generalized measurement error model above. Consider,

$$Y_p = \sum_{m=1}^{M} X_m^* \mathbf{I}(S_{pm}) + \varepsilon_p \qquad\qquad p = 1, \ldots, P$$

Y_p are observed bids of bidder $1, \ldots, P$, X_m^*, $m = 1, \ldots, M$ are unobserved "common" values, $\mathbf{I}(S_{pm})$ is an indicator that bidder p's valuation includes the common value X_m^* and ε_p, $p = 1, \ldots, P$ are unobserved private values.

Gautier and Kitamura (2009) apply results from deconvolution to nonparametrically estimate the density in a random coefficients binary choice model. This paper may offer an opportunity to extend their results.

3 Identification

The following theorem is the main result of this paper. It provides sufficient conditions for identification of U_{m^*} when some of the unobservables are linearly and statistically dependent. The proof is constructive so that sample analogs can be used for estimation. The unobservables and observables can be discrete or continuous.

Theorem 1. *If Assumptions 1,2 and 3 are satisfied then the distribution of U_{m^*} is identified.*

Start with the system of equations in Assumption 1 and left multiply $Y = AU$ by B from Assumption 3i. to get a new system of equations. The objective of left multiplying by B is to take linear combinations of $Y_1, \ldots, Y_P$ so that U_{m^*} is in as many equations as possible and the other unobservables are in as few equations as possible. Each equation with U_{m^*} provides an opportunity to identify its distribution. The sparsity makes it easier to distinguish the effects of U_{m^*} from other unobservables. BA should be a reduced row echelon matrix with U_{m^*} as a free variable (in many economic models A is already in reduced row echelon form).

For the purposes of proving identification assume that B is the identity matrix ($B = I_P$) so that $\widetilde{A} = A$ otherwise left multiply $Y = AU$ by B, relabel the variables and work with the new system of equations.

I now show that the linear relationship $Y = AU$ will be retained after a transformation to log characteristic functions. The characteristic function of Y is

$$\begin{aligned}\phi_{Y_1,\ldots,Y_P}(t_1,\ldots,t_P) &= E\left[\exp\left(i(a_{11}t_1+\ldots+a_{P1}t_P)Z_1+\ldots+i(a_{1M}t_1+\ldots+a_{PM}t_P)W_{M_{dep}}\right)\right] \\ &= \prod_{m=1}^{M_{ind}}\phi_{Z_m}\left(\sum_{p=1}^{P}a_{pm}t_p\right)E\left[\exp\left(iW_1\sum_{p=1}^{P}a_{pM_{ind}+1}t_p+\ldots+iW_{M_{dep}}\sum_{p=1}^{P}a_{pM}t_p\right)\right]\end{aligned}$$

where the first equality follows from the definition of the characteristic function and the second equality follows from the dependence structure in Assumption 2. Take the natural logarithm of both sides and let $\varphi_Y(t) = \ln\phi_Y(t)$, $\varphi_m(t) = \ln\phi_{Z_m}(t)$, $m = 1,\ldots,M_{ind}$ and

$$\varphi_W(\omega) = \varphi_{W_1,\ldots,W_{M_{dep}}}(\omega_1,\ldots,\omega_{M_{dep}}) = \ln E\left[\exp\left(iW_1\omega_1+\ldots+iW_{M_{dep}}\omega_{M_{dep}}\right)\right]$$

where $Y = (Y_1,\ldots,Y_P)$, $W = (W_1,\ldots,W_{M_{dep}})$ and $t = (t_1,\ldots,t_p)$ then

$$\varphi_Y(t) = \sum_{m=1}^{M_{ind}}\varphi_m\left(\sum_{p=1}^{P}a_{pm}t_p\right)+\varphi_W\left(\sum_{p=1}^{P}a_{pM_{ind}+1}t_p,\ldots,\sum_{p=1}^{P}a_{pM}t_p\right)$$

Take the derivative with respect to $t_1,\ldots t_P$

$$\begin{pmatrix}\frac{\partial\varphi_Y(t)}{\partial t_1}\\ \vdots \\ \frac{\partial\varphi_Y(t)}{\partial t_P}\end{pmatrix} = \begin{pmatrix}a_{11} & \ldots & a_{1M}\\ \vdots & \ddots & \vdots\\ a_{P1} & \ldots & a_{PM}\end{pmatrix}\begin{pmatrix}\varphi_1'\left(\sum_{p=1}^{P}a_{p1}t_p\right)\\ \vdots \\ \varphi_{M_{ind}}'\left(\sum_{p=1}^{P}a_{pM_{ind}}t_p\right)\\ \frac{\partial\varphi_W\left(\sum_{p=1}^{P}a_{pM_{ind}+1}t_p,\ldots,\sum_{p=1}^{P}a_{pM}t_p\right)}{\partial\omega_1}\\ \vdots \\ \frac{\partial\varphi_W\left(\sum_{p=1}^{P}a_{pM_{ind}+1}t_p,\ldots,\sum_{p=1}^{P}a_{pM}t_p\right)}{\partial\omega_{M_{dep}}}\end{pmatrix}$$

Observe that the new system of equations is identical to (1) except random variables are replaced by first-order partial derivatives of characteristic functions. Dependent unobservables cannot be separated so remain together as part of nonseparable multidimensional functions.

By Assumption 3 equation p^* will be used to identify U_{m^*}

$$\frac{\partial\varphi_Y(t)}{\partial t_{p^*}} = \sum_{m=1}^{M_{ind}}a_{p^*m}\varphi_m'\left(\sum_{p=1}^{P}a_{pm}t_p\right)+\sum_{m=1}^{M_{dep}}a_{p^*M_{ind}+m}\frac{\partial\varphi_W\left(\sum_{p=1}^{P}a_{pM_{ind}+1}t_p,\ldots,\sum_{p=1}^{P}a_{pM}t_p\right)}{\partial\omega_m}$$

$$= \sum_{\substack{m=1 \\ a_{p^*m}\neq 0}}^{M_{ind}} a_{p^*m}\varphi'_m\left(\sum_{p=1}^{P} a_{pm}t_p\right) + i \sum_{\substack{m=1 \\ a_{p^*M_{ind}+m}\neq 0}}^{M_{dep}} a_{p^*M_{ind}+m} \frac{E\left[W_m \exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'} \sum_{p=1}^{P} a_{pM_{ind}+m'}t_p\right)\right]}{E\left[\exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'} \sum_{p=1}^{P} a_{pM_{ind}+m'}t_p\right)\right]}$$

$$= \sum_{a_{p^*m}\neq 0} a_{p^*m}\varphi'_m\left(\mathbf{I}\left(a_{p^*m}\neq 0\right)\sum_{p=1}^{P} a_{pm}t_p\right)$$

$$+ i \sum_{a_{p^*M_{ind}+m}\neq 0} a_{p^*M_{ind}+m} E\left[W_m \exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'}\mathbf{I}\left(U_{m^*}\in W \text{ or } a_{p^*M_{ind}+m'}\neq 0\right)\sum_{p=1}^{P} a_{pM_{ind}+m'}t_p\right.\right.$$

$$\left.\left.+i\sum_{m'=1}^{M_{dep}} W_{m'}\mathbf{I}\left(U_{m^*}\notin W \text{ and } a_{p^*M_{ind}+m'}=0\right)\sum_{p=1}^{P} a_{pM_{ind}+m'}t_p\right)\right] \Big/ E\left[\exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'}\sum_{p=1}^{P} a_{pM_{ind}+m'}t_p\right)\right]$$

$$= \sum_{a_{p^*m}\neq 0} a_{p^*m}\varphi'_m\left(\mathbf{I}\left(a_{p^*m}\neq 0\right)A'_m t\right)$$

$$+ i \sum_{a_{p^*M_{ind}+m}\neq 0} a_{p^*M_{ind}+m} E\left[W_m \exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'}\mathbf{I}\left(U_{m^*}\in W \text{ or } a_{p^*M_{ind}+m'}\neq 0\right)A'_{M_{ind}+m'}t\right.\right.$$

$$\left.\left.+i\sum_{m'=1}^{M_{dep}} W_{m'}\mathbf{I}\left(U_{m^*}\notin W \text{ and } a_{p^*M_{ind}+m'}=0\right)A'_{M_{ind}+m'}t\right)\right] \Big/ E\left[\exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'}A'_{M_{ind}+m'}t\right)\right]$$

$$= \sum_{a_{p^*m}\neq 0} a_{p^*m}\varphi'_m\left(A_m^{U,p^*,m^*}t\right)$$

$$+ i \sum_{a_{p^*M_{ind}+m}\neq 0} a_{p^*M_{ind}+m} E\left[W_m \exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'}A^{U,p^*,m^*}_{M_{ind}+m'}t\right.\right.$$

$$\left.\left.+i\sum_{m'=1}^{M_{dep}} W_{m'}\mathbf{I}\left(U_{m^*}\notin W \text{ and } a_{p^*M_{ind}+m'}=0\right)A'_{M_{ind}+m'}t\right)\right] \Big/ E\left[\exp\left(i\sum_{m'=1}^{M_{dep}} W_{m'}A'_{M_{ind}+m'}t\right)\right]$$

where the third equality holds because $\mathbf{I}\left(U_{m^*}\in W \text{ or } a_{p^*M_{ind}+m'}\neq 0\right) + \mathbf{I}\left(U_{m^*}\notin W \text{ and } a_{p^*M_{ind}+m'}=0\right) \equiv 1$ and the last equality follows from the definition of A^{U,p^*,m^*}.

Consider two cases:

1. $U_{m^*}\in W$. Hence, $\mathbf{I}\left(U_{m^*}\notin W \text{ and } a_{p^*M_{ind}+m'}=0\right) = 0$ and $\mathbf{I}\left(U_{m^*}\in W \text{ or } a_{p^*M_{ind}+m'}\neq 0\right) = 1$. Let $t = se^{m^*}$ from assumption 3ii. then

$$\frac{\partial\varphi_Y(t)}{\partial t_{p^*}} = \sum_{a_{p^*m}\neq 0} a_{p^*m}\varphi'_m(0) + i\sum_{a_{p^*M_{ind}+m}\neq 0} a_{p^*M_{ind}+m}\frac{E\left[W_m\exp\left(iU_{m^*}s\right)\right]}{E\left[\exp\left(iU_{m^*}s\right)\right]}$$

$$
\begin{aligned}
&= i \sum_{a_{p^*m} \neq 0} a_{p^*m} E\left[Z_m\right] + i \sum_{a_{p^*M_{ind}+m} \neq 0} a_{p^*M_{ind}+m} \frac{E\left[E\left[W_m | U_{m^*}\right] \exp\left(iU_{m^*}s\right)\right]}{E\left[\exp\left(iU_{m^*}s\right)\right]} \\
&= i \sum_{a_{p^*m} \neq 0} a_{p^*m} E\left[Z_m\right] + i \sum_{\substack{a_{p^*M_{ind}+m} \neq 0 \\ M_{ind}+m \neq m^*}} a_{p^*M_{ind}+m} E\left[W_m\right] \frac{E\left[\exp\left(iU_{m^*}s\right)\right]}{E\left[\exp\left(iU_{m^*}s\right)\right]} + a_{p^*m^*} \frac{iE\left[U_{m^*} \exp\left(iU_{m^*}s\right)\right]}{E\left[\exp\left(iU_{m^*}s\right)\right]} \\
&= i \sum_{a_{p^*m} \neq 0} a_{p^*m} E\left[Z_m\right] + i \sum_{\substack{a_{p^*M_{ind}+m} \neq 0 \\ M_{ind}+m \neq m^*}} a_{p^*M_{ind}+m} E\left[W_m\right] + a_{p^*m^*} \frac{iE\left[U_{m^*} \exp\left(iU_{m^*}s\right)\right]}{E\left[\exp\left(iU_{m^*}s\right)\right]} \\
&= i \sum_{a_{p^*m} \neq 0} a_{p^*m} E\left[Z_m\right] + i \sum_{\substack{a_{p^*M_{ind}+m} \neq 0 \\ M_{ind}+m \neq m^*}} a_{p^*M_{ind}+m} E\left[W_m\right] + a_{p^*m^*} \frac{\partial \ln E\left[\exp\left(iU_{m^*}s\right)\right]}{\partial s} \\
&= i \sum_{a_{p^*m} \neq 0} a_{p^*m} E\left[Z_m\right] + i \sum_{\substack{a_{p^*M_{ind}+m} \neq 0 \\ M_{ind}+m \neq m^*}} a_{p^*M_{ind}+m} E\left[W_m\right] + a_{p^*m^*} \varphi'_{m^*}(s)
\end{aligned}
$$

where the first equality follows from the choice of t and Assumption 3ii., the second equality follows from $\varphi'_m(0) = \frac{\phi'_m(0)}{\phi_m(0)} = iE[Z_m]$ and the third equality follows from Assumption 3iii. $E\left[W_m | U_{m^*}\right] = E\left[W_m\right]$.

By the second fundamental theorem of calculus

$$
\varphi_{m^*}(s) = \varphi_{m^*}(0) + \int_0^s \varphi'_{m^*}(u)\mathrm{d}u = \frac{1}{a_{p^*m^*}} \left(\int_0^s \frac{\partial \varphi_Y(ue^{m^*})}{\partial t_{p^*}} \mathrm{d}u - is \sum_{m=1}^{M_{ind}} a_{p^*m} E[Z_m] - is \sum_{\substack{m=1 \\ M_{ind}+m \neq m^*}}^{M_{dep}} a_{p^*M_{ind}+m} E[W_m
$$

$$
\phi_{m^*}(s) = \exp\left(\frac{1}{a_{p^*m^*}} \left(\int_0^s \frac{iE\left[Y_{p^*} \exp(iu \sum_{p=1}^P e_p^{m^*} Y_p)\right]}{\phi_Y(ue^{m^*})} \mathrm{d}u - is \sum_{m=1}^{M_{ind}} a_{p^*m} E[Z_m] - is \sum_{\substack{m=1 \\ M_{ind}+m \neq m^*}}^{M_{dep}} a_{p^*M_{ind}+m} E[W_m] \right) \right.
$$

where I used

$$
\frac{\partial \varphi_Y(t)}{\partial t_{p^*}} = \frac{\partial \phi_Y(t)/\partial t_{m^*}}{\phi_Y(t)} = \frac{iE\left[Y_{p^*} \exp(i \sum_{p=1}^P t_p Y_p)\right]}{\phi_Y(t)}
$$

2. $U_{m^*} \notin W$. Hence, $\mathbf{I}\left(U_{m^*} \notin W \text{ and } a_{p^*M_{ind}+m'} = 0\right) = \mathbf{I}\left(a_{p^*M_{ind}+m'} = 0\right)$ and $\mathbf{I}\left(U_{m^*} \in W \text{ or } a_{p^*M_{ind}+m'} \neq 0\right) = \mathbf{I}\left(a_{p^*M_{ind}+m'} \neq 0\right)$. Let $t = se^{m^*}$ from assumption 3ii. then

$$
\begin{aligned}
\frac{\partial \varphi_Y(t)}{\partial t_{p^*}} &= a_{p^*m^*} \varphi'_{m^*}(s) + \sum_{\substack{a_{p^*m} \neq 0 \\ m \neq m^*}} a_{p^*m} \varphi'_m(0) \\
&+ i \sum_{a_{p^*M_{ind}+m} \neq 0} a_{p^*M_{ind}+m} \frac{E\left[W_m \exp\left(i \sum_{m'=1}^{M_{dep}} W_{m'} \mathbf{I}\left(a_{p^*M_{ind}+m'} = 0\right) \sum_{p=1}^P a_{pM_{ind}+m'} e_p^{m^*}\right)\right]}{E\left[\exp\left(i \sum_{m'=1}^{M_{dep}} W_{m'} \mathbf{I}\left(a_{p^*M_{ind}+m'} = 0\right) \sum_{p=1}^P a_{pM_{ind}+m'} e_p^{m^*}\right)\right]}
\end{aligned}
$$

$$= a_{p^*m^*}\varphi'_{m^*}(s) + i \sum_{\substack{a_{p^*m}\neq 0\\ m\neq m^*}} a_{p^*m} E[Z_m]$$

$$+ i \sum_{a_{p^*M_{ind}+m}\neq 0} a_{p^*M_{ind}+m} \frac{E\left[E\left[W_m | W_{-p^*-m}\right] \exp\left(i \sum_{a_{p^*M_{ind}+m'}=0} W_{m'} \sum_{p=1}^{P} a_{pM_{ind}+m'} e_p^{m^*}\right)\right]}{E\left[\exp\left(i \sum_{a_{p^*M_{ind}+m'}=0} W_{m'} \sum_{p=1}^{P} a_{pM_{ind}+m'} e_p^{m^*}\right)\right]}$$

$$= a_{p^*m^*}\varphi'_{m^*}(s) + i \sum_{\substack{a_{p^*m}\neq 0\\ m\neq m^*}} a_{p^*m} E[Z_m] + i \sum_{a_{p^*M_{ind}+m}\neq 0} a_{p^*M_{ind}+m} E\left[W_m\right]$$

where the first equality follows from the choice of t and Assumption 3ii., the second equality follows from $\varphi'_m(0) = \frac{\phi'_m(0)}{\phi_m(0)} = iE[Z_m]$ and the third equality follows from Assumption 3iii. $E\left[W_m | W_{p^*m}\right] = E\left[W_m\right]$.

By the second fundamental theorem of calculus

$$\varphi_{m^*}(s) = \varphi_{m^*}(0) + \int_0^s \varphi'_{m^*}(u)\mathrm{d}u = \frac{1}{a_{p^*m^*}} \left(\int_0^s \frac{\partial \varphi_Y(ue^{m^*})}{\partial t_{p^*}} \mathrm{d}u - is \sum_{\substack{m=1\\ m\neq m^*}}^{M_{ind}} a_{p^*m} E[Z_m] - is \sum_{m=1}^{M_{dep}} a_{p^*M_{ind}+m} E[W_m] \right)$$

$$\phi_{m^*}(s) = \exp\left(\frac{1}{a_{p^*m^*}} \left(\int_0^s \frac{iE\left[Y_{p^*} \exp(iu \sum_{p=1}^{P} e_p^{m^*} Y_p)\right]}{\phi_Y(ue^{m^*})} \mathrm{d}u - is \sum_{\substack{m=1\\ m\neq m^*}}^{M_{ind}} a_{p^*m} E[Z_m] - is \sum_{m=1}^{M_{dep}} a_{p^*M_{ind}+m} E[W_m] \right)\right)$$

where I used

$$\frac{\partial \varphi_Y(t)}{\partial t_{p^*}} = \frac{\partial \phi_Y(t)/\partial t_{m^*}}{\phi_Y(t)} = \frac{iE\left[Y_{p^*} \exp(i \sum_{p=1}^{P} t_p Y_p)\right]}{\phi_Y(t)}$$

Regardless of the dependence structure of U and choices of B, p^* and e^{m^*}, the expression for $\phi_{m^*}(s)$ takes the form

$$\phi_{m^*}(s) = \exp\left(i \left(\int_0^s \frac{\sum_j C_j E\left[Y_j \exp(iu \sum_{p=1}^{P} e_p^{m^*} Y_p)\right]}{\phi_Y(ue^{m^*})} \mathrm{d}u - s \sum_{\substack{m=1\\ m\neq m^*}}^{M} c_{p^*m} E[Z_m] - s \sum_{\substack{m=1\\ m\neq m^*}}^{M_{dep}} c_{p^*M_{ind}+m} E[W_m] \right)\right)$$

where C_j and c_{p^*m}, $m = 1, \ldots, M$ are constants.

$\phi_{m^*}(s)$ is not identified if the absolute value of the argument inside the exponential is ∞. Notice that $|\exp(iu)| \leq 1$ for $u \in \mathcal{R}$ so the reason that $\phi_{m^*}(s)$ is not identified is because as $|u| \to \infty$, $\exp(iu) = \cos(u) + i\sin(u)$ oscillates very rapidly.

Assumption 3iii., $E[|U_m|] < \infty$, $m = 1, \ldots, M$ is too strong in some cases. A weaker condition is $E[|U_m|] < \infty$ when $c_{p^*m} \neq 0$ and $\sum_j |C_j| E\left[|Y_j|\right] = \sum_{j'} C'_{j'} E\left[|U_{j'}|\right] < \infty$.

Using Assumption 3iii. and $|\exp(iu)| \leq 1$ for $u \in \mathcal{R}$

$$
\begin{aligned}
|\phi_{m^*}(s)| &\leq \exp\left(\int_0^s \left| \frac{i \sum_j C_j E\left[Y_j \exp(iu \sum_{p=1}^P e_p^{m^*} Y_p)\right]}{\phi_Y(ue^{m^*})} \right| \mathrm{d}u + |s| \sum_{\substack{m=1 \\ m \neq m^*}}^{M} |c_{p^*m}| E[|Z_m|] + |s| \sum_{\substack{m=1 \\ m \neq m^*}}^{M_{dep}} |c_{p^* M_{ind}+m}| E[|W_m|] \right) \\
&\leq \exp\left(\widetilde{C}_1 \int_0^s \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u + |s| \widetilde{C}_2 \right)
\end{aligned}
$$

for some positive finite constants $\widetilde{C}_1$ and $\widetilde{C}_2$. This final expression is finite if and only if $\int_0^s \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u < \infty$. A common assumption in the literature is that $|\phi_{U_m}(u)| \neq 0$ for all u in their support and all unobservables. Because integration over u allows ϕ_{U_m} to take arbitrary values on sets of zero Lebesgue measure and have no effect on the integral, control over every single point in the space is excessive. Furthermore, $\phi_Y(ue^{m^*}) = \phi_{Y'e^{m^*}}(u) = E\left[\exp(iuY'e^{m^*})\right] = E\left[\exp(iu(AU)'e^{m^*})\right] = \phi_{(AU)'e^{m^*}}(u)$ so only the characteristic function of $(AU)'e^{m^*}$ needs to be restricted. Evdokimov and White (2011) notice these weaker restrictions and allow some of the unobservables to have isolated zeros and others to have no zero restrictions. Another possibility is to assume that $\int_{S_{m^*}} \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u < \infty$ where S_{m^*} is the support of ϕ_{m^*} (this is an absolute-integrability condition).

I now bound $|\phi_{m^*}(s)|$ using Assumption 3iv., which includes isolated zeros and the absolute-integrability condition as special cases. By Assumption 3iv., if $\int_X |\phi_{Y'e^{m^*}}(s)| \mathrm{d}s = 0$ then $\int_X |\phi_{m^*}(s)| \mathrm{d}s = 0$ for all sets X of nonzero Lebesgue measure. Assume $|\phi_Y(0)| = 0$ and there are no sets X between 0 and s where $\int_X |\phi_{Y'e^{m^*}}(s)| \mathrm{d}s = 0$.[17] By Assumption 3iv.

$$
|\phi_{m^*}(s)| \leq \exp\left(\widetilde{C}_1 \int_0^s \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u + |s| \widetilde{C}_2 \right) < \infty
$$

Finally, there is a bijection between the density and characteristic function of U_{m^*}. The density of U_{m^*} for all u in the support of U_{m^*} is identified by the inverse Fourier transform,

$$
f_{m^*}(u) = \frac{1}{2\pi} \int e^{-isu} \phi_{m^*}(s) \mathrm{d}s
$$

The unobservables are identified sequentially so sometimes identification of nuisance random variables can be avoided. Other times U_{m^*} will not be identified from Theorem 1 but another unobservable $U_{m^{**}}$, $m^{**} \neq m^*$ is identified. If $U_{m^{**}}$ satisfies some independence conditions then $U_{m^{**}}$ is treated as observed, included as part of

[17] If there are sets X between 0 and s then the lower limit of integration needs to change.

Y and U_{m^*} is identified from a smaller system of equations. Thus, identification can be achieved by recursively using Theorem 1 and moving identified unobservables that satisfy some independence conditions to the left hand side of the equation. Example 1 Estimator B, Example 1 Estimator C and Example 3 use this recursive technique for identification.

ϕ_{m^*} is overidentified if B, p^* or the solution e^{m^*} are not unique (note that B and p^* can be different for different m^*). In addition to the expressions for identification in this paper there are several other identification strategies which lead to even more ways to express the unobservable distributions. For example, Bonhomme and Robin (2010) identify unobservables using the second-order partial derivatives of the observable characteristic functions (which leads to an infinite number of solutions). Carrasco and Florens (2010) base their estimator of an unobservable distribution on the spectral decomposition of the convolution operator.

Therefore, I provide an algorithm for the practitioner to choose B, p^* and e^{m^*} and identify ϕ_{m^*} using Theorem 1:

1. Choose B so that BA is in reduced row echelon form with U_{m^*} as one of the free variables.
2. Check Assumption 3 for $p^* = 1, \ldots, P$ and choose $e^{m^*} = A^{p^* m^* +} e_{m^*}$ where $A^{p^* m^* +}$ is the Moore-Penrose pseudoinverse.[18]
3. An estimator is then based on the expressions in Theorem 1.

I now return to the three empirical illustrations from the Model and Assumptions section and identify the unobservables.

Example 1: Measurement Error With Two Measurements (Continued)

Applying B, p^* and e^{m^*} from the Model and Assumptions section and Theorem 1, the characteristic functions of the unobservables are expressed as functionals of observable first order partial derivatives of log characteristic functions:

$$\text{Estimator A:} \;\; \phi_{X^*}(s) = \exp\left(\int_0^s \frac{iE\left[Y_1 \exp(iuY_2)\right]}{\phi_{Y_2}(u)} \mathrm{d}u - isE\left[\varepsilon_1\right] \right)$$
$$\text{Estimator B:} \;\; \phi_{X^*}(s) = \frac{\phi_{Y_1}(s)}{\phi_{\varepsilon_1}(s)} \text{ where } \phi_{\varepsilon_1}(t) = \exp\left(\int_0^s \frac{iE\left[Y_1 \exp(iu(Y_1 - Y_2))\right]}{\phi_{Y_1 - Y_2}(u)} \mathrm{d}u - isE\left[X^*\right] \right)$$
$$\text{Estimator C:} \;\; \phi_{X^*}(s) = \frac{\phi_{Y_1}(s)}{\phi_{\varepsilon_1}(s)} \text{ where } \phi_{\varepsilon_1}(t) = \exp\left(\int_0^s \frac{iE\left[(Y_1 - Y_2)\exp(iuY_1)\right]}{\phi_{Y_1}(u)} \mathrm{d}u + isE\left[\varepsilon_2\right] \right)$$

All the estimators are similar in spirit to the expression in Kotlarski (1967). In the economics literature, Estimator A is used by Li and Vuong (1998) and Cunha, et al. (2010), Estimator B is used by Evdokimov

[18] The Moore-Penrose pseudoinverse is available as a built-in function in MATLAB.

(2011) and Estimator C I have not seen in the economics literature.

The three estimators convey different intuition for identification of the distribution of X^*. In Estimator A, the term inside the integration, $\frac{iE[Y_1\exp(iuY_2)]}{\phi_{Y_2}(u)} = \left.\frac{\partial \ln\phi_Y}{\partial t_1}\right|_{(0,u)}$, is the partial derivative of the log characteristic function of Y with respect to the first argument evaluated at $(0,u)$. The partial derivative with respect to the first argument leads to small changes in the log characteristic functions of the random variables in the equation $Y_1 = X^* + \varepsilon_1$, which are $\ln\phi_{Y_1}$, $\ln\phi_{X^*}$ and $\ln\phi_{\varepsilon_1}$. The small change in $\ln\phi_{X^*}$ leads to a small change in $\ln\phi_{Y_2}$ because $Y_2 = X^* + \varepsilon_2$. The small change in $\ln\phi_{Y_2}$ is determined by evaluation at $(0,u)$ (and is only caused by $\ln\phi_{X^*}$ because ε_2 is independent of (X^*,ε_1)) and identifies the derivative of $\ln\phi_{X^*}$ and in turn the distribution of X^*. Loosely speaking, one "moves" the first equation and observes the second equation. The "movement" in the first equation causes X^* and ε_1 to move. X^* leads to "movement" in the second equation (that is not caused by ε_2), so by observing the second equation X^* is identified.

In Estimator B, ε_1 is identified by "moving" $Y_1 = X^* + \varepsilon_1$ and observing $Y_1 - Y_2 = \varepsilon_1 - \varepsilon_2$, which "moves" only because of ε_1. In Estimator C, ε_1 is identified by "moving" $Y_3 = Y_1 - Y_2 = \varepsilon_1 - \varepsilon_2$, and observing $Y_1 = X^* + \varepsilon_1$, which "moves" only because of ε_1. In both estimators X^* is then identified from $Y_1 = X^* + \varepsilon_1$ using deconvolution.

These are not the only possible expressions for ϕ_{X^*}. For example, $\phi_{X^*}(t) = \exp\left(\int_0^t \left(\frac{iE[Y_2\exp(iuY_1)]}{\phi_{Y_1}(u)}\right) \mathrm{d}u - itE\left[\varepsilon_2\right]\right)$ is also a solution.[19]. Bonhomme and Robin (2010) prove that for any e that satisfies $Y'e = e_1Y_1 + e_2Y_2 = 1$, ϕ_{X^*} is identified by

$$\phi_{X^*}(s) = \exp\left(\int_0^s \left(-\frac{iE\left[Y_1Y_2\exp(iuY'e)\right]}{\phi_{Y'e}(u)} + \frac{E\left[Y_1\exp(iuY'e)\right]}{\phi_{Y'e}(u)}\frac{E\left[Y_2\exp(iuY'e)\right]}{\phi_{Y'e}(u)}\right) \mathrm{d}u + isE\left[X^*\right]\right)$$

It is not known which expression is the best but one objective of this paper is to use finite sample simulations to compare Estimator A, Estimator B, Estimator C, the Bonhomme and Robin (2010) estimator with their choice of $e = (\frac{1}{2}, \frac{1}{2})$ (labeled Estimator D) and an estimator based on deconvolution when ε_1 is known (labeled Estimator E).

Estimator A is the only consistent estimator under the weaker conditions $E[\varepsilon_1]$ known, ε_2 independent of (X^*,ε_1) and $\phi_{Y_2}(u) = 0$ on a set of zero Lebesgue measure.

More generally, it is not known what is the best estimator in the measurement error model with $P \geq 2$ measurements of the unknown variable X^*

$$X_p = X^* + \varepsilon_p, \quad p = 1, \ldots, P$$

19 $\frac{iE[Y_2\exp(iuY_1)]}{\phi_{Y_1}(u)}$ is the derivative of $\ln\phi_Y(u)$ with respect to the second argument evaluated at $(u,0)$ Hence, identification comes from "moving" $Y_2 = X^* + \varepsilon_2$ and observing $Y_1 = X^* + \varepsilon_1$, which "moves" only because of X^*

$$X^*, \varepsilon_1, \ldots \varepsilon_P \text{ mutually independent}$$

In practice all but two of the observations are ignored. This seems like a waste of information. A solution of $\phi_{X^*}(s)$ that uses all the observations is

$$\phi_{X^*}(s) = \exp\left(\int_0^s \frac{iE\left[Y_1 \exp\left(iu\frac{1}{P-1}\sum_{p=2}^P Y_p\right)\right]}{\phi_{\left(\frac{1}{P-1}\sum_{p=2}^P Y_p\right)}(u)} \mathrm{d}u - isE\left[\varepsilon_1\right]\right)$$

Example 2: Earnings Dynamics (Continued)

Applying B, p^* and e^{m^*} from the Model and Assumptions section and Theorem 1, the characteristic functions of the unobservables are expressed as functionals of observable first order partial derivatives of log characteristic functions:

$$\begin{aligned}
\phi_{\eta_2}(s) &= \exp\left(-i\int_0^s \frac{E\left[Y_2 \exp\left(iuY_1\right)\right]}{\phi_{Y_1}(u)} \mathrm{d}u\right) \\
\phi_{\eta_3}(s) &= \exp\left(i\int_0^s \frac{E\left[Y_2 \exp\left(-iuY_3\right)\right]}{\phi_{Y_3}(-u)} \mathrm{d}u\right) \\
\phi_{\varepsilon_2}(s) &= \exp\left(i\int_0^s \left(\frac{E\left[Y_1 \exp\left(iu(Y_1)\right)\right]}{\phi_{Y_1}(u)} + \frac{E\left[Y_2 \exp\left(iu(Y_1)\right)\right]}{\phi_{Y_1}(u)} + \frac{E\left[Y_3 \exp\left(iu(Y_1)\right)\right]}{\phi_{Y_1}(u)}\right) \mathrm{d}u\right) \\
\phi_{\varepsilon_3}(s) &= \exp\left(-i\int_0^s \frac{E\left[Y_2 \exp\left(iu(Y_1+Y_2+Y_3)\right)\right]}{\phi_{Y_1+Y_2+Y_3}(u)} \mathrm{d}u\right) \\
\phi_{\varepsilon_4}(s) &= \exp\left(i\int_0^s \left(\frac{E\left[Y_1 \exp\left(iu(Y_2)\right)\right]}{\phi_{Y_2}(u)} + \frac{E\left[Y_2 \exp\left(iu(Y_2)\right)\right]}{\phi_{Y_2}(u)} + \frac{E\left[Y_3 \exp\left(iu(Y_2)\right)\right]}{\phi_{Y_2}(u)}\right) \mathrm{d}u\right)
\end{aligned}$$

[20] Most of the literature on earnings dynamics only identifies and estimates the variances of the persistent and transitory shocks (see Hsiao (1986)). Horowitz and Markatou (1996) identify the distribution of an earnings dynamics model without persistent shocks ($\varepsilon_{nt} = 0$) and independent identically distributed transitory shocks (η_{nt}) using characteristic functions (deconvolution formulas). Bonhomme and Robin (2011) identify mutually independent unobservables using second order partial derivatives of observable log characteristic functions.

[20] Explicit identification of η_2, for example, follows from

$$\begin{aligned}
\frac{\partial \ln \phi_Y(u,0,0)}{\partial t_2} &= \frac{E\left[i\left(-\eta_2+\eta_3\right)\exp\left(iu\eta_2\right)\right]}{E\left[\exp\left(iu\eta_2\right)\right]} + \frac{E\left[i\varepsilon_3 \exp\left(iu\varepsilon_2\right)\right]}{E\left[\exp\left(iu\varepsilon_2\right)\right]} \\
&= -\frac{E\left[i\eta_2 \exp\left(iu\eta_2\right)\right]}{E\left[\exp\left(iu\eta_2\right)\right]} + \frac{E\left[i\exp\left(iu\eta_2\right)E\left[\eta_3|\eta_2\right]\right]}{E\left[\exp\left(u\eta_2\right)\right]} + \frac{E\left[i\exp\left(iu\varepsilon_2\right)E\left[\varepsilon_3|\varepsilon_2\right]\right]}{E\left[\exp\left(iu\varepsilon_2\right)\right]} \\
&= -\frac{E\left[i\eta_2 \exp\left(iu\eta_2\right)\right]}{E\left[\exp\left(iu\eta_2\right)\right]} \\
&= -\frac{\partial \ln \phi_{\eta_2}(u)}{\partial u}
\end{aligned}$$

Example 3: Production Function (Continued)

Identification proceeds in three steps: First identify a and b, second identify the distribution of unobservables and third identify the functions g and h.

For any $l, \bar{l}$ ($l \neq \bar{l}$) in the support of L_{ijt} and any $k, \bar{k}$ in the support of K_{ijt}, a is identified by

$$a = \frac{E\left[Y_{ij2}|L_{ij2} = l, K_{ij2} = \bar{k}\right] - E\left[Y_{ij2}|L_{ij2} = \bar{l}, K_{ij2} = \bar{k}\right]}{E\left[Y_{ij1}|L_{ij1} = l, K_{ij1} = k\right] - E\left[Y_{ij1}|L_{ij1} = \bar{l}, K_{ij1} = k\right]}$$

b is similarly identified.

Let $X := (L_{ij1}, L_{i'j'1}, L_{ij'2}, L_{i'j2}, K_{ij1}, K_{i'j'1}, K_{ij'2}, K_{i'j2}) = (l, \bar{l}, l, \bar{l}, k, \bar{k}, \bar{k}, k) =: x$, and apply B, p^* and e^{m^*} from the Model and Assumptions section and Theorem 1. The conditional characteristic functions of the unobservables are expressed as functionals of the first order partial derivatives of log characteristic functions:

$$\phi_{\varepsilon_{ij1}}(s|X = x) = \exp\left(i \int_0^s \frac{E\left[(abY_{ij1} - bY_{ij'2} + b^2 Y_{i'j'1} - aY_{i'j'2}) \exp\left(iu(Y_{ij1})\right)\right]}{ab\phi_{Y_{ij1}}(u)} \mathrm{d}u\right)$$

$$\phi_{\varepsilon_{ij'2}}(s|X = x) = \exp\left(i \int_0^s \frac{E\left[-abY_{ij1} + bY_{ij'2} - b^2 Y_{i'j'1} + aY_{i'j'2} \exp\left(iu(Y_{ij'2})\right)\right]}{b\phi_{Y_{ij'2}}(u)} \mathrm{d}u\right)$$

$$\phi_{\varepsilon_{i'j'1}}(s|X = x) = \exp\left(i \int_0^s \frac{E\left[b^2 Y_{ij1} - aY_{ij'2} + abY_{i'j'1} - bY_{i'j'2} \exp\left(iu(Y_{i'j'1})\right)\right]}{ab\phi_{Y_{i'j'1}}(u)} \mathrm{d}u\right)$$

$$\phi_{\varepsilon_{i'j2}}(s|X = x) = \exp\left(i \int_0^s \frac{E\left[-b^2 Y_{ij1} + aY_{ij'2} - abY_{i'j'1} + bY_{i'j'2} \exp\left(iu(Y_{i'j'2})\right)\right]}{b\phi_{Y_{i'j'2}}(u)} \mathrm{d}u\right)$$

$$\phi_{g(l,\alpha_i)}(s|X = x) = \left(\phi_{\varepsilon_{ij1}}(s)\right)^{\frac{b^2}{a^2-b^2}} \exp\left(i \int_0^s \frac{E\left[(-b^2 Y_{ij1} + aY_{ij'2} - abY_{i'j'1} + bY_{i'j'2}) \exp\left(iuY_{ij1}\right)\right]}{(a^2 - b^2)\phi_{Y_{ij1}}(u)} \mathrm{d}u\right)$$

$$\phi_{g\left(\bar{l},\alpha_{i'}\right)}(s|X = x) = \left(\phi_{\varepsilon_{ij'2}}(s)\right)^{\frac{b}{b^2-a^2}} \exp\left(i \int_0^s \frac{E\left[(-abY_{ij1} + bY_{ij'2} - b^2 Y_{i'j'1} + aY_{i'j'2}) \exp\left(iuY_{ij'2}\right)\right]}{(a^2 - b^2)\phi_{Y_{ij'2}}(u)} \mathrm{d}u\right)$$

$$\phi_{h(k,\beta_j)}(s|X = x) = \left(\phi_{\varepsilon_{i'j'1}}(s)\right)^{\frac{ab}{b^2-a^2}} \exp\left(i \int_0^s \frac{E\left[(a^2 Y_{ij1} - aY_{ij'2} + abY_{i'j'1} - bY_{i'j'2}) \exp\left(iuY_{i'j'1}\right)\right]}{(a^2 - b^2)\phi_{\frac{Y_{i'j'1}}{(}u)}} \mathrm{d}u\right)$$

$$\phi_{h\left(\bar{k},\beta_{j'}\right)}(s|X = x) = \left(\phi_{\varepsilon_{i'j2}}(s)\right)^{\frac{a}{a^2-b^2}} \exp\left(i \int_0^s \frac{E\left[(abY_{ij1} - bY_{ij'2} + a^2 Y_{i'j'1} - aY_{i'j'2}) \exp\left(iuY_{i'j'2}\right)\right]}{(a^2 - b^2)\phi_{Y_{i'j'2}}(u)} \mathrm{d}u\right)$$

The distributions are identified by the inverse Fourier transform.

The next step is to identify the functions g and h. Consider first a random effects model and identification of g. Assume α_i is independent of everything else, α_i is uniformly distributed on $[0, 1]$ and $g(l, \alpha)$ is increasing in α for all l in the support of L_{ijt}. I now use a result from Matzkin (2003). Let $Q_{g(l,\alpha_i)|X=x}(\alpha|X = x)$ be the conditional quantile function (identified from the distributions). For all l in the support of L_{ijt} and $\alpha \in (0,1)$

$$g(l,\alpha) = g(l, Q_{\alpha_i|X=x}(\alpha|X=x)) = Q_{g(l,\alpha_i)|X=x}(\alpha|X=x)$$

where the first equality follows from the random effects assumption that α_i is uniform and independent of everything else and the second equality follows from the assumption that g is increasing in α. Similar assumptions and arguments identify h.

Consider now a fixed effects model and identification of g. Assume $g(l,\alpha)$ is increasing in α for all l in the support of L_{ijt}, $g(\bar{l},\alpha) = \alpha$ for all α and α is continuous. Suppose the same labor-individual works with two capital-individuals over two periods and let $X := (L_{ij1}, L_{ij'1}, L_{ij'2}, L_{ij2}, K_{ij1}, K_{ij'1}, K_{ij'2}, K_{ij2}) = (l, \bar{l}, l, \bar{l}, k, \bar{k}, \bar{k}, k) =: \bar{x}$

$$\begin{aligned}
Y_{ij1} &= g\left(l, \alpha_i\right) + h\left(k, \beta_j\right) + \varepsilon_{ij1} \\
Y_{ij'2} &= ag\left(\bar{l}, \alpha_i\right) + bh\left(k, \beta_j\right) + \varepsilon_{ij'2} \\
Y_{i'j'1} &= g\left(\bar{l}, \alpha_i\right) + h\left(\bar{k}, \beta_{j'}\right) + \varepsilon_{i'j'1} \\
Y_{i'j2} &= ag\left(l, \alpha_i\right) + bh\left(\bar{k}, \beta_{j'}\right) + \varepsilon_{i'j2}
\end{aligned}$$

then

$$\begin{aligned}
\left(\frac{1}{b^2-a^2}\right)\left(b^2 Y_{ij1} - aY_{ij'2} + abY_{i'j'1} - bY_{i'j'2}\right) &= g\left(l,\alpha_i\right) + \left(\frac{1}{b^2-a^2}\right)\left(b^2\varepsilon_{ij1} - a\varepsilon_{ij'2} + ab\varepsilon_{i'j'1} - b\varepsilon_{i'j2}\right) \\
\left(\frac{1}{b^2-a^2}\right)\left(abY_{ij1} - bY_{ij'2} + b^2 Y_{i'j'1} - aY_{i'j'2}\right) &= g\left(\bar{l},\alpha_i\right) + \left(\frac{1}{b^2-a^2}\right)\left(ab\varepsilon_{ij1} - b\varepsilon_{ij'2} + b^2\varepsilon_{i'j'1} - a\varepsilon_{i'j2}\right)
\end{aligned}$$

Take the log characteristic functions on both sides and solve for $\phi_{g(l,\alpha_i)}$ and $\phi_{g\left(\bar{l},\alpha_i\right)} = \phi_{\alpha_i}$ (the equality follows from the assumption that $g(\bar{l},\alpha) = \alpha$ for all α)

$$\begin{aligned}
\phi_{g(l,\alpha_i)}(s|X=\bar{x}) &= \frac{\phi_{\left(\frac{b^2 Y_{ij1} - aY_{ij'2} + abY_{i'j'1} - bY_{i'j'2}}{b^2-a^2}\right)}(s|X=\bar{x})}{\phi_{\left(\frac{b^2\varepsilon_{ij1}}{b^2-a^2}\right)}(s|X=x)\,\phi_{\left(\frac{-a\varepsilon_{ij'2}}{b^2-a^2}\right)}(s|X=x)\,\phi_{\left(\frac{ab\varepsilon_{i'j'1}}{b^2-a^2}\right)}(s|X=x)\,\phi_{\left(\frac{-b\varepsilon_{i'j2}}{b^2-a^2}\right)}(s|X=x)} \\
\phi_{\alpha_i}(s|X=\bar{x}) &= \frac{\phi_{\left(\frac{abY_{ij1} - bY_{ij'2} + b^2 Y_{i'j'1} - aY_{i'j'2}}{b^2-a^2}\right)}(s|X=\bar{x})}{\phi_{\left(\frac{ab\varepsilon_{ij1}}{b^2-a^2}\right)}(s|X=\bar{x})\,\phi_{\left(\frac{-b\varepsilon_{ij'2}}{b^2-a^2}\right)}(s|X=\bar{x})\,\phi_{\left(\frac{b^2\varepsilon_{i'j'1}}{b^2-a^2}\right)}(s|X=\bar{x})\,\phi_{\left(\frac{-a\varepsilon_{i'j2}}{b^2-a^2}\right)}(s|X=\bar{x})}
\end{aligned}$$

where both equalities follow from mutual independence of α_i, ε_{ij1}, $\varepsilon_{ij'2}$, $\varepsilon_{i'j'1}$ and $\varepsilon_{i'j2}$. The characteristic functions on the right hand side were identified in the second step so the characteristic functions and distributions of $g(l, \alpha_i)$ and α_i are also identified. For all l in the support of L_{ijt} and $\alpha \in (0,1)$, the structural function $g(l, \alpha)$ is now identified by

$$g(l,\alpha) = g(l, Q_{\alpha_i|X=\bar{x}}(F_{\alpha_i|X=\bar{x}}(\alpha|X=\bar{x})|X=\bar{x})) = Q_{g(l,\alpha_i)|X=\bar{x}}(F_{\alpha_i|X=\bar{x}}(\alpha|X=\bar{x})|X=\bar{x})$$

where the first equality follows because Q is the inverse of F and α is continuous and the second equality follows from the assumption that g is increasing in α. Similar assumptions and arguments identify h.[21]

Evdokimov (2011) considers the model $Y_{it} = m(X_{it}, \alpha_i) + \varepsilon_{it}$ and identifies m from a panel data with $T = 2$ and $X_{i1} = X_{i2} = x$. His novel interpretation of Kotlarski's result is one of the inspirations for this paper. The production function example is a generalization of Evdokimov (2011) with multiple unobserved heterogeneity. The second and third step are almost the same as Evdokimov's.

4 Estimation

The proof of Theorem 1 is constructive so sample analogs are used to replace unknown population quantities. Given i.i.d observations $\{Y_1, \dots, Y_N\}$ where $Y_n = (Y_{n1}, \dots, Y_{nP})$, estimate characteristic functions $\phi_Y(t) = E[\exp(iY't)]$ by

$$\widehat{\phi}_Y(t) = E_N[\exp(iY't)] = \frac{1}{N}\sum_{n=1}^{N} \exp(iY_n't)$$

and estimate first-order partial derivatives of characteristic functions with respect to the p^{th} argument $\frac{\partial \phi_Y(t)}{\partial t_p} = iE[Y_p \exp(it'Y)]$ by

$$\widehat{\phi}_{Yp}(t) = \widehat{\frac{\partial \phi_Y(t)}{\partial t_p}} = iE_N[Y_p \exp(iY't)] = \frac{i}{N}\sum_{n=1}^{N} Y_{np}\exp(iY_n't)$$

All the characteristic functions take the form

$$\phi_{m^*}(s) = \exp\left(i\int_0^s \frac{\sum_j C_j E\left[Y_j \exp(iuY'e^{m^*})\right]}{E\left[\exp(iuY'e^{m^*})\right]} du - is\sum_{\substack{m=1\\ m\neq m^*}}^{M} c_{p^*m}E[Z_m] - is\sum_{\substack{m=1\\ m\neq m^*}}^{M_{dep}} c_{p^*M_{ind}+m}E[W_m] \right)$$

$E[U_m]$, $m = 1, \dots, M$, is known so assume $E[U_m] = 0$. Each term in the summation of $\sum_j C_j E\left[Y_j \exp(iuY'e^{m^*})\right]$ is dealt with in the same way so to avoid unnecessary complicating the notation assume $C_{p^*} = 1$ and $C_p = 0$

[21] An alternative proof uses $h(\bar{k}, \beta) = 0$ for all β and follows Evdokimov (2011).

when $p \neq p^*$. The characteristic function

$$\phi_{m^*}(s) = \exp\left(i \int_0^s \frac{E\left[Y_{p^*} \exp(iuY'e^{m^*})\right]}{E\left[\exp\left(iuY'e^{m^*}\right)\right]} \mathrm{d}u \right)$$

is estimated by

$$\widehat{\phi}_{m^*}(s) = \exp\left(i \int_0^s \frac{E_N\left[Y_{p^*} \exp\left(uY'e^{m^*}\right)\right]}{E_N\left[\exp\left(iuY'e^{m^*}\right)\right]} \mathrm{d}u \right)$$

The density of U_{m^*} is obtained by inverting the characteristic function using the inverse Fourier transformation,

$$f_{m^*}(u) = \frac{1}{2\pi} \int e^{-isu} \phi_{m^*}(s) \mathrm{d}s$$

This integral does not converge when the characteristic function is replaced by its sample analog so the integral is truncated on a compact interval $[-S_N, S_N]$ with $S_N \to \infty$ as $N \to \infty$. The density of U_{m^*} is estimated by

$$\widehat{f}_{m^*}(u) = \frac{1}{2\pi} \int e^{-itu} \widehat{\phi}_{m^*}(s) \phi_K\left(sh_N\right) \mathrm{d}s$$

where $\phi_K(s) = \int \exp(isu)H(u)\mathrm{d}u$ is the Fourier transform of a kernel K supported on $[-1, 1]$ and $h_N = \frac{1}{S_N}$ is the bandwidth of the kernel. The kernel leads to relatively slow convergence rates but solves any irregularity problems by smoothing the estimator. I use the second order kernel

$$K(u) = \frac{48\cos(u)}{\pi u^4}\left(1 - \frac{15}{u^2}\right) - \frac{144\sin(u)}{\pi u^5}\left(2 - \frac{5}{u^2}\right)$$

whose Fourier transform is

$$\phi_K(s) = (1 - s^2)^3 \mathbf{I}(s \in [-1, 1])$$

This kernel is often used in the deconvolution literature (see Delaigle and Gijbels (2002)).

Computing $\widehat{f}_{m^*}(u)$ requires two integrations: one integration to recover the ϕ_{m^*} and another integration to perform the inverse Fourier transform. When ϕ_Y is small there is rapid oscillation. To keep the estimation stable I use the trapezoid rule for integration (which is slower but more robust than some other approximate integration techniques) and divide the interval into $1,000$ grid points of integration (this is particularly important when the characteristic function has isolated zeros). The choice of bandwidth is a topic that I do not investigate in this paper but the simulations suggest that this is an important topic of research. When the bandwidth is too large the confidence intervals are wide because convergence is unstable when ϕ_Y is small, which causes ϕ_{m^*} to oscillate rapidly and hence inaccurate estimation. When the bandwidth is too small the estimated density is far from the theoretical density because too much of the tail of ϕ_{m^*} is ignored. For every choice of an underlying theoretical density, the median estimator gave very good approximations so there was no need to adjust the estimator to deal with instances when the characteristic function had countably many zeros. Nevertheless, the confidence bands around the median were large when characteristic functions were small. This can be dealt with by having bands around zero that approach zero as $N \to \infty$ as in Hu and Ridder (2010).

5 Asymptotic Theory

In this section, I study the asymptotic properties of the estimator $\widehat{f}_{m^*}$. I will show that $\widehat{f}_{m^*}$ is a uniformly consistent estimator of f_{m^*}. The proofs are in the Appendix and are similar to Bonhomme and Robin (2010).

Characterizing the asymptotic properties of $\widehat{f}_{m^*}$ proceeds in three steps. Lemma 1 and Lemma 2 bound

$$\sup_{t\in[-T,T]^P} \left|E_N\left[Y_p \exp\left(iY't\right)\right] - E\left[Y_p \exp\left(iY't\right)\right]\right| \quad \text{and} \quad \sup_{t\in[-T,T]^P} \left|E_N\left[\exp\left(iY't\right)\right] - E\left[\exp\left(iY't\right)\right]\right|$$

Theorem 1 uses Lemma 1 and Lemma 2 to bound

$$\sup_{s\in[-T,T]} \left|\widehat{\phi}_{m^*}(s) - \phi_{m^*}(s)\right|$$

and Theorem 2 uses Theorem 1 to bound

$$\sup_{u} \left|\widehat{f}_{m^*}(u) - f_{m^*}(u)\right|$$

These upper bounds provide sufficient conditions for consistency and help understand which estimators have fastest uniform convergence rates.

Lemma 1. *Let F denote the cumulative distribution function of Y and F_N the empirical cumulative distribution function corresponding to a sample $(Y^1, \ldots, Y^N)$ of N independent identically distributed draws from F. Assume $E[|Y_p|^2] < \infty$, $p = 1, \ldots, P$. Let*

$$\begin{aligned} T_N &= C_1 N^{\delta/2} \qquad C_1, \delta > 0 \\ \varepsilon_N &= C_2 \frac{\ln N}{\sqrt{N}} \end{aligned}$$

where $C_2^2 > 64(2 + P(1 + \delta))$ *then*

$$\sup\left|\widehat{\frac{\partial \phi_Y(t)}{\partial t_p}} - \frac{\partial \phi_Y(t)}{\partial t_p}\right| < \varepsilon_N \qquad a.s.$$

when N *tends to infinity.*[22]

As $N \to \infty$, Lemma 1 bounds the estimation error on ϕ_{Yp} on the compact interval $[-T_N, T_N]^P$ by $\mathcal{O}\left(\ln N/\sqrt{N}\right)$ provided that T_N does not grow faster than some power of N.[23]

The strategy in the proof is standard for finding uniform convergence rates in the empirical processes literature. First, restrict the problem to a compact space that expands as N increases. A "divide and conquer" strategy is applied by dividing the problem into two components: one component where the second moment is larger than M_N and one component where the second moment is smaller than M_N (M_N grows at an appropriate rate). When the second moment is larger than M_N, use a Chernoff bound to show that the probability of this event goes to zero. When the second moment is smaller than M_N, first use the Heine-Borel theorem (which states that any open cover of a compact space has a finite subcover) to cover the space by a finite number of arbitrarily small balls. At the center of each ball, bound the probability that the distance between $\widehat{\phi}_{Yp}$ and ϕ_{Yp} (the estimation error of ϕ_{Yp}) is bigger than any $\varepsilon > 0$ using an exponential-type bound. Because every point is arbitrarily close to the center of one of the balls, this implies that the estimation error at every point is bounded. The last step is to use the Borel-Cantelli lemma to show that the event that the estimation error is large happens only a finite number of times so for large enough N the estimation error is small almost surely.

Lemma 2. *Let*

$$T_N = \widetilde{C}_1 \left(\frac{\ln N}{N}\right)^{\frac{1}{2}} N^{\frac{\widetilde{\delta}}{2}} \qquad\qquad \widetilde{C}_1 > 0, \widetilde{\delta} > 1$$

$$\varepsilon_N = \widetilde{C}_2 \left(\frac{\ln N}{N}\right)^{\frac{1}{2}}$$

where $\widetilde{C}_2^2 > 64(2 + P\widetilde{\delta})$ *then*

$$\sup\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right| < \varepsilon_N \qquad a.s.$$

when N *tends to infinity.*

As $N \to \infty$, Lemma 2 bounds the estimation error on the compact interval $[-T_N, T_N]^P$ by $\mathcal{O}\left(\left(\frac{\ln N}{N}\right)^{\frac{1}{2}}\right)$

[22] To simplify notation I suppress the subscript $t \in [-T_N, T_N]^P$ in $\sup_{t\in[-T_N,T_N]^P}$ unless there is some ambiguity or the sup is not over this region.

[23] $Z_N = \mathcal{O}(a_N)$ means that there exists $C > 0$ such that $Z_N \leq C a_N$.

provided that T_N does not grow too quickly. The rate of convergence is the same as Horowitz and Markatou (1996).[24] The rate of convergence is faster than Lemma 1 but the choices of T_N and ε_N from Lemma 1 also lead to uniform convergence to $\phi_Y(t)$.

The main difference between the proof of Lemma 1 and Lemma 2, is that in the proof of Lemma 2 I do not control the size of the second moment (or any moment because the argument in the expectation $|\exp(iY't)| \leq 1$). Hence, there is no divide and conquer strategy and the proof starts by using the Heine-Borel theorem and from there proceeds identically to the proof of Lemma 1.

When the support of $\phi_Y(t)$ is bounded, Li and Vuong (1998) bound $\sup_t \left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|$ by $\mathcal{O}\left((\ln\ln N/N)^{\frac{1}{2}}\right)$.[25] The slower convergence rates of Lemma 1 and Lemma 2 are because ε_N cannot shrink too quickly relative to the growing compact space $[-T_N, T_N]^P$.

When the support of $\phi_Y(t)$ is unbounded, Bonhomme and Robin (2010) bound

$$\sup\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|,\ \sup\left|\widehat{\frac{\partial\phi_Y(t)}{\partial t_p}} - \frac{\partial\phi_Y(t)}{\partial t_p}\right| \text{ and } \sup\left|\widehat{\frac{\partial^2\phi_Y(t)}{\partial t_{p_1}\partial t_{p_2}}} - \frac{\partial^2\phi_Y(t)}{\partial t_{p_1}\partial t_{p_2}}\right|$$

with the same bounds as Lemma 1. An important assumptive difference is no moment restrictions when bounding the uniform convergence rate of $\widehat{\phi}_Y(t)$, $E[|Y_p^2|]$ finite when bounding the uniform convergence rate of $\widehat{\frac{\partial\phi_Y(t)}{\partial t_p}}$ and $E[|Y_p^4|]$ finite when bounding the uniform convergence rate of $\widehat{\frac{\partial^2\phi_Y(t)}{\partial t_p^2}}$.

Theorem 2. *Define $(A)^+ = \max\{A, 0\}$. Choose ε_N and T_N according to Lemma 1 then there exists $C_7 > 0$ such that*

$$\sup_{s\in[-S_N,S_N]}\left|\widehat{\phi}_{m^*}(s) - \phi_{m^*}(s)\right| \leq C_7\varepsilon_N\int_{-S_N}^{S_N}\frac{1}{|\phi_Y(ue^*)|\left(|\phi_Y(ue^*)| - \varepsilon_N\right)^+}\mathrm{d}u$$

where for consistency $\varepsilon_N\int_{-S_N}^{S_N}\frac{1}{|\phi_Y(ue^{m^*})|\left(|\phi_Y(ue^{m^*})|-\varepsilon_N\right)^+}\mathrm{d}u$ goes to zero as $N\to\infty$.[26]

The rate of uniform convergence to $\phi_{m^*}(s)$ depends on S_N, ϕ_Y and the relative sizes of $|\phi_Y|$ and ε_N. The main concern in estimation is similar to the concern in identification; large $\int_0^s \frac{1}{\widehat{\phi}_Y(ue^*)}\mathrm{d}u$ or equivalently $\widehat{\phi}_Y$ small on sets of nonzero Lebesgue measure. This leads to rapid and large oscillations in the estimator that makes it hard to get accurate estimations.

For large enough N, small ϕ_Y implies small $\widehat{\phi}_Y$. Hence, convergence rates are slower when ϕ_Y is small over sets of nonzero Lebesgue measure. In a seminal paper, Fan (1991) distinguishes two classes of distributions with different convergence rates: if ϕ_Y has no zeros but decays at an exponential rate then ϕ_Y has thin tails and f_Y

[24]Hu and Ridder (2010) noticed that Horowitz and Markatou (1996) use Pollard (1984) without taking into account a correction for unbounded support. I obtain the same convergence rates with this correction.

[25]For bounded support a proof can be based on Csörgo (1981).

[26]Sometimes the faster convergence rates in Lemma 2 apply.

is called super-smooth. If ϕ_Y has no zeros but decays at a polynomial rate then ϕ_Y has fat tails and f_Y is called ordinary-smooth. ϕ_{m^*} is harder to estimate when f_Y is super-smooth because ϕ_Y approaches zero faster.

When $|\phi_Y|$ is smaller than ε_N then the ε_N neighborhood around ϕ_Y includes zero. Uniform convergence is a "worst case" type of convergence rate so if $\widehat{\phi}_Y$ can be zero it will be. Hence, if $|\phi_Y| < \varepsilon_N$ on sets of nonzero Lebesgue measure then the estimator will not converge uniformly.

Hu and Ridder (2010) solve the problem of $|\phi_Y| < \varepsilon_N$ by restricting $|\widehat{\phi}| > \eta_N$, with $\eta_N > 0$ converging to zero. In the Monte Carlo Simulations section I did not modify the estimator and had no issues when the characteristic function of a distribution (e.g. uniform) had countably many zeros. This is possibly because uniform convergence is too strong a criteria for convergence.

S_N effects the uniform convergence rate in two ways. Integration causes errors to accumulate so as S_N increases the convergence rate decreases.[27] As S_N increases in a neighborhood around 0 or becomes large, ϕ_Y shrinks and the convergence rate decreases.

Importantly, the uniform convergence rate also depends on the shape of ϕ_{m^*} itself. If ϕ_{m^*}, for example, has bounded support then even if ϕ_Y is unbounded, the values of ϕ_Y outside the support of ϕ_{m^*} have no effect (because of the limits of integration). In general, a super-smooth f_{m^*} and an ordinary-smooth f_Y leads faster convergence rates. Super-smooth f_{m^*} means that the tails of ϕ_{m^*} are thin so most the mass of ϕ_{m^*} is near the origin so that the weights on the basis functions $\exp(-isu)$ become small very quickly and have little impact on the estimate of f_{m^*}. Furthermore, any information in the tail of ϕ_Y is relatively larger than ϕ_{m^*} because ϕ_Y has fatter tails than ϕ_{m^*}.

I compare the uniform convergence rate of Theorem 1 to some other estimators in the deconvolution literature. There are three criteria that effect convergence rates and distinguish estimators: 1. the support of ϕ_{X^*} (bounded or unbounded) 2. the functions used in the estimator (estimators can be functionals of characteristic functions, first order partial derivatives of characteristic function, second-order partial derivatives of characteristic functions or other spectral decompositions of the convolution operator) 3. smoothness of f_Y and smoothness of f_{m^*} (Recently, the literature also makes a distinction between vanishing and nonvanishing characteristic functions but I think the speed at which ϕ_Y and ϕ_{m^*} approach zero is the dominant factor affecting convergence).

Li and Vuong (1998) and Bonhomme and Robin (2010) consider the measurement error model with two measurements and ϕ_{X^*} and ϕ_{ε_p}, $p = 1, 2$ ordinary-smooth (this implies that $\phi_Y(s) = s^{-\beta}$, $\beta > 1$). Li and Vuong (1998) assume that ϕ_{X^*} has bounded support and their estimator is a functional of (nonvanishing) characteristic functions. They obtain a uniform convergence rate of $\mathcal{O}\left(\left(\frac{N}{\ln \ln N}\right)^{-\frac{1}{2}+\alpha}\right)$, $0 < \alpha < \frac{1}{2}$. Bonhomme

[27]Notice the difference between T_N and S_N. $[-T_N, T_N]^P$ is the growing compact support of the P-dimensional variable t, which is the domain of ϕ_Y while $[-S_N, S_N]$ is the growing compact support of s, which is the domain of ϕ_{m^*}.

and Robin (2010) assume that ϕ_{X^*} has unbounded support and their estimator is a functional of second-order partial derivatives of the characteristic functions. They obtain a uniform convergence rate of $\mathcal{O}\left(\frac{\ln N}{N^{\frac{1}{2}-\left(1+\frac{3}{2}\beta\right)\delta}}\right)$. By Theorem 1, I obtain a uniform convergence rate of $\mathcal{O}\left(\frac{\ln N}{N^{\frac{1}{2}-\left(\frac{1}{2}+\beta\right)\delta}}\right)$ which is faster than Bonhomme and Robin (2010) because I use ϕ_{Yp} rather than $\frac{\partial^2\phi_Y}{\partial p_1 \partial p_2}$ but slower than Li and Vuong (1998) because of their bounded support assumption.

Theorem 3. *Choose ε_N and T_N according to Lemma 1 and assume the convergence rate from Theorem 1 applies then there exists C_8 such that*

$$\begin{aligned}
&\sup_u \left|\widehat{f}_{m^*}(u) - f_{m^*}(u)\right| \\
&\leq \frac{S_N C_7 \varepsilon_N}{\pi} \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ve^*)|\left(|\phi_Y(ve^*)| - \varepsilon_N\right)^+} \mathrm{d}v + \frac{1}{2\pi} \sup_{s\in[-1,1]} |m(s)|\, h_N^q \int_{-S_N}^{S_N} |s|^q |\phi_{m^*}(s)| \mathrm{d}s \\
&+ \frac{1}{2\pi}\int_{-\infty}^{-S_N} |\phi_{m^*}(s)|\mathrm{d}s + \frac{1}{2\pi}\int_{S_N}^{\infty} |\phi_{m^*}(s|)\mathrm{d}s
\end{aligned}$$

where for consistency $\varepsilon_N \int_{-S_N}^{S_N} \frac{1}{\left|\phi_Y(ve^{m^*})\right|\left(\left|\phi_Y(ve^{m^*})\right|-\varepsilon_N\right)^+}\mathrm{d}v$ goes to zero as $N \to \infty$.

The rate of uniform convergence of $\widehat{f}_{m^*}(u)$ depends on four components: the characteristic function of observables ϕ_Y (and the relative size of ϕ_Y and ε_N), the Fourier transform of the kernel ϕ_K, the characteristic function of the unobservable ϕ_{m^*} and S_N. The first component is the estimation error of ϕ_{m^*} and the last three terms are the loss because of the inversion.

The effect of ϕ_Y (and the relative size of ϕ_Y and ε_N) is through the first term, which is the estimation error of ϕ_{m^*}. The effects are thus the same as in Theorem 1: small ϕ_Y makes the first term big and leads to a slower convergence rate.

The second term results from the smoothing kernel. This term eventually approaches zero. The choice of bandwidth, is an important choice parameter but beyond the scope of this paper.

The other components have an ambiguous effect on the uniform convergence rate of $\widehat{f}_{m^*}(u)$ because of the way they interact with other components in each term. The tension between large and small S_N arises because large S_N makes the last two terms (the tails) smaller but the first term (estimation error for ϕ_{m^*}) and second term (the inverse of the smoothing kernel) bigger.

If f_{m^*} is super-smooth then the last two terms are smaller because ϕ_{m^*} decays quickly but the first term is bigger because ϕ_{m^*} is harder to estimate and the second term is bigger because ϕ_{m^*} is relatively big near the origin.

Many of the estimators in the literature, like the estimator in this paper, integrate over the inverse of ϕ_Y. Thus, a heuristic argument for an estimator with the fastest uniform convergence rate will make $\phi_Y(s)$ large where $\phi_{m^*}(s)$ is large because $\phi_Y(s)$ is the marginal information for $\phi_{m^*}(s)$. The decay of the inverse Fourier transform due to e^{-isu} also suggests that accurate estimates of $\phi_{m^*}(s)$ near the origin are more important because they get more weight.

In the deconvolution literature, Carroll and Hall (1988) and Fan (1991) consider the model $Y = X^* + \varepsilon$ where the distribution of ε is known. They obtain logarithmic rates of convergence $\mathcal{O}\left((\ln N)^{-2(m+\alpha-l)/\beta}\right)$ when f_{X^*} is super smooth and f_Y is ordinary smooth.

Bonhomme and Robin (2010) consider the measurement error model with two measurements where f_{X^*} and f_Y are ordinary smooth($\phi_Y(s) = s^{-\beta}$, $\beta > 1$ and $\phi_{m^*}(s) = s^{-\alpha}$, $\alpha > 1$). To simplify the problem they assume the kernel that corresponds to $\phi_K(s) = \mathbf{I}(s \in [-1,1])$ and obtain the uniform convergence rate $\mathcal{O}\left(\frac{\ln N}{N^{\frac{1}{2}-\frac{3}{2}(1+\beta)\delta}} + \frac{1}{N^{\delta}2(\alpha-1)}\right)$. By Theorem 2, the uniform convergence is $\mathcal{O}\left(\frac{\ln N}{N^{\frac{1}{2}-(1+\beta)\delta}} + \frac{1}{N^{\delta}2(\alpha-1)}\right)$, which is faster than Bonhomme and Robin (2010).

More generally, Bonhomme and Robin (2010) consider the same multi-factor model as in this paper and obtain a uniform rate of convergence of $\frac{C_1 S_N^3 \varepsilon_N}{g(S_N)^3} + \frac{C_2}{S_N^q 2\pi}\int_{-S_N}^{S_N} |s|^q |\phi_{m^*}(s)| \mathrm{d}s + 2\int_{S_N}^{\infty} |\phi_{m^*}(s)| \mathrm{d}s$. At first this might seem slower than other convergence rates but notice that the first term is very small for values of S_N near the origin. If enough of the distribution is identified in the neighborhood $[-1,1]$ (where $\frac{S_N^3}{g(S_N)^3}$ is small) then their estimator converges very quickly. Thus super-smooth distributions will probably converge fastest using their estimator. The Monte Carlo simulations in the next section confirm this as the convergence rates are fastest especially when the unobservables are normally distributed.

6 Monte Carlo Simulations

This section presents a Monte Carlo study of the finite sample properties of five estimators of the density of X^* in the measurement error model with repeated measurements:

$$X_1 = X^* + \varepsilon_1$$
$$X_2 = X^* + \varepsilon_2$$

The data is generated from one of the following specifications of the distributions of X^*, ε_1 and ε_2

Experiment	f_{X^*}	f_{ε_1}	f_{ε_2}
1	Norm(0,1)	Norm(0,1)	Norm(0,1)
2	Gamma(2,1)	Norm(0,1)	Norm(0,1)
3	Gamma(5,1)	Norm(0,1)	Norm(0,1)
4	Lognormal	Norm(0,1)	Norm(0,1)
5	$\frac{400}{403}N(0,\frac{1}{2})+\frac{3}{403}N(0,\frac{406}{6})$	Norm(0,1)	Norm(0,1)
6	$\frac{1}{2}N(-2,1)+\frac{1}{2}N(2,1)$	Norm(0,1)	Norm(0,1)
7	Norm(0,1)	Norm(0,x^{*2})	Norm(0,1)
8	Unif(0,2)	0	0

where x^{*2} (the variance of ε_1 in Experiment 7) is the square of the value that is attained by the random variable X^* in each trial (hence X^* and ε_1 are dependent). I compare five estimators:

	Estimator
A	$\widehat{\phi}_{X^*}(s)=\exp\left(\int_0^s \frac{iE_N[Y_1\exp(iuY_2)]}{E_N[\exp(iuY_2)]}du-isE\left[\varepsilon_1\right]\right)$
B	$\widehat{\phi}_{X^*}(s)=\frac{\phi_{Y_1}(s)}{\widehat{\phi}_{\varepsilon_1}(s)}$ where $\widehat{\phi}_{\varepsilon_1}(t)=\exp\left(\int_0^s \frac{iE_N[Y_1\exp(iu(Y_1-Y_2))]}{E_N[\exp(iu(Y_1-Y_2))]}du-isE\left[X^*\right]\right)$
C	$\widehat{\phi}_{X^*}(s)=\frac{\phi_{Y_1}(s)}{\widehat{\phi}_{\varepsilon_1}(s)}$ where $\widehat{\phi}_{\varepsilon_1}(t)=\exp\left(\int_0^s \frac{iE_N[(Y_1-Y_2)\exp(iuY_1)]}{E_N[\exp(iuY_1)]}du+isE\left[\varepsilon_2\right]\right)$
D	$\widehat{\phi}_{X^*}(s)=\exp\left(\int_0^s\left(-\frac{iE_N\left[Y_1Y_2\exp\left(\frac{iu}{2}(Y_1+Y_2)\right)\right]}{E_N\left[\exp\left(\frac{iu}{2}(Y_1+Y_2)\right)\right]}+\frac{E_N\left[Y_1\exp\left(\frac{iu}{2}(Y_1+Y_2)\right)\right]}{E_N\left[\exp\left(\frac{iu}{2}(Y_1+Y_2)\right)\right]}\frac{E_N\left[Y_2\exp\left(\frac{iu}{2}(Y_1+Y_2)\right)\right]}{E_N\left[\exp\left(\frac{iu}{2}(Y_1+Y_2)\right)\right]}\right)du+isE\left[X^*\right]\right)$
E	$\widehat{\phi}_{X^*}(s)=\frac{\phi_{Y_1}(s)}{\phi_{\varepsilon_1}(s)}$

where the first estimator is used by Li and Vuong (1998), the second estimator is used by Evdokimov (2011), the third estimator is new, the fourth estimator is used by Bonhomme and Robin (2010) and the fifth estimator is used by Hu and Ridder (2005).

Only $Y_1,\dots,Y_N$ is observed for Estimators A, B, C and D. $Y_1,\dots,Y_N$ is observed and the distribution of ε_1 is known for Estimator E. Assume $E[\varepsilon_1]$ is known for Estimator A. Assume $E[X^*]$ is known for Estimators B and D and assume $E[\varepsilon_2]$ is known for Estimator C.

I generate 100 simulations of sample size $N=100$, $N=1,000$ and $N=10,000$. The grid on the x-axis is divided into $1,000$ equidistant grid points for integration in both characteristic function space and density space.

The results are summarized graphically in Figures 1 to 8. Figure 1 reports the outcomes of 100 simulations of sample size $1,000$ where the data is generated according to Experiment 1. The first column, is an estimate of the real part of ϕ_{X^*}, the second column is an estimate of the imaginary part of ϕ_{X^*} and the third column is an estimate of f_{X^*}. On each graph the thin solid line represents population quantities, the thick solid line represents the median of the simulations and the dashed lines represent the 10-90% pointwise confidence bands. The first row depicts the results of Estimator A, the second row depicts the results of Estimator B, the third row depicts the results of Estimator C, the fourth row depicts the results of Estimator D and the fifth row depicts the results of Estimator E. Figures 2 to 8 are the same as Figure 1 except for Experiments 2 to 8.

To provide an indication of relative finite sample efficiencies of the estimators, Tables 1, 2 and 3 report the mean integrated squared error (MISE) of each estimator for $N = 100$, $N = 1,000$ and $N = 10,000$ respectively where

$$\text{MISE} = E\left[\int \left(\widehat{f}_{X^*}(x) - f_{X^*}(x)\right)^2 dx\right]$$

Overall the median estimators do very well. Estimators B and C are robust to isolated zeros (Experiment 8). As expected, only Estimator A is consistent in Experiment 7 (due to dependence structure of unobservables). Estimators B, C and E perform well when the distribution is bimodal (Experiment 6) while Estimator A converges slowly and Estimator D does not converge. The confidence bands are wide when ϕ_{X^*} is close to zero for extended intervals as in Experiments 2 and 5 for example. The experiment with slowest convergence rates is Experiment 4 (Lognormal) but notice that $\widehat{\phi}_{m^*}$ approximates ϕ_{m^*} very well as can be seen from the first two columns in Figure 4. The problem is in the inversion and even with the population characteristic function I need very large S_N for convergence. The characteristic function of the Lognormal distribution has fat tails that impact the estimate of f_{X^*} but are hard to estimate accurately due to accumulation of estimation error and although the tails are fat they are still relatively small ($0.05 \leq |\phi_{X^*}(s)| \leq 0.12$ for $2 \leq |s| \leq 12$). Surprisingly, when the distribution of ε_1 is known, convergence is not fastest. Estimator E performs poorly in Experiments 2, 3 and 4 when the distributions are asymmetric (Gamma and Lognormal) and in Experiment 8 when the distribution is discontinuous (Uniform). Estimator D had the fastest convergence rates in Experiments 1 (Standard Normal), 2 (Gamma(2,1)) and 3 (Gamma(5,1)). Estimator D did not converge in Experiment 6 because of bimodality and Experiment 8 because of discontinuity (and is biased in Experiment 7). Estimator C converges the fastest in experiments 6 (Bimodal), which is possibly because the choice of the matrix B creates a third equation $Y_3 = Y_1 - Y_2$, which perhaps averages errors. Estimator C probably also converges fastest in Experiments 5 (Unimodal) and 8 (Uniform).

7 Conclusions

This paper presented techniques for nonparametrically identifying and estimating a short panel data model with many unobservables and with some unobservables arbitrarily dependent. The paper derived the uniform rate of convergence and presented a Monte Carlo study which suggests that the estimators are robust and perform well in finite samples.

The performance of the estimators depended critically on the shape of the unobservable and observable distributions. The large number of available estimators makes the study of systematically identifying and estimating the best estimator very interesting. This paper (and several others) suggests that zeros of the operator are a major stumbling block and efficiency may come from trying to avoid them. An extension of this paper is to model the equations as a known multivariate nonlinear function of unobservables. The choice of bandwidth was not considered in this paper but the simulations suggest this is an important area to investigate.

Appendix

7.1 Proof of Identification from Bonhomme and Robin (2010)

Bonhomme and Robin (2010) assume the setup of Assumption 1, $U_1, \ldots, U_M$ mutually independent, $E[U_m] = 0$ for $m = 1, \ldots, M$ and a rank condition on a matrix related to A that will be defined later.

The characteristic function of Y is

$$\begin{aligned}\phi_{Y_1,\ldots,Y_P}(t_1, \ldots, t_P) &= E\left[\exp i\left((a_{11}t_1 + \ldots + a_{P1}t_P)U_1 + \ldots + (a_{1M}t_1 + \ldots + a_{PM}t_P)U_M\right)\right] \\ &= \prod_{m=1}^{M} \phi_{U_m}\left(\sum_{p=1}^{P} a_{pm}t_p\right)\end{aligned}$$

where the first equality follows from the definition of the characteristic function and the second equality follows from mutual independence. Take the natural logarithm of both sides and let $\varphi_Y(t) = \ln \phi_Y(t)$, $\varphi_m(t) = \ln \phi_{Z_m}(t)$, $m = 1, \ldots, M_{ind}$. The first order partial derivatives are

$$\begin{pmatrix} \frac{\partial \varphi_Y(t)}{\partial t_1} \\ \vdots \\ \frac{\partial \varphi_Y(t)}{\partial t_P} \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & a_{1M} \\ \vdots & \ddots & \vdots \\ a_{P1} & \ldots & a_{PM} \end{pmatrix} \begin{pmatrix} \varphi_1'\left(\sum_{p=1}^{P} a_{p1}t_p\right) \\ \vdots \\ \varphi_M'\left(\sum_{p=1}^{P} a_{pM}t_p\right) \end{pmatrix}$$

The second order partial derivatives are

$$\begin{pmatrix} \frac{\partial^2 \varphi_Y(t)}{\partial t_1^2} \\ \vdots \\ \frac{\partial^2 \varphi_Y(t)}{\partial t_{p_1}\partial t_{p_2}} \\ \vdots \\ \frac{\partial^2 \varphi_Y(t)}{\partial t_P^2} \end{pmatrix} = \begin{pmatrix} a_{11}^2 & \ldots & a_{1M}^2 \\ \vdots & \ddots & \vdots \\ a_{p_1 1}a_{p_2 1} & \ldots & a_{p_1 M}a_{p_2 M} \\ \vdots & \ddots & \vdots \\ a_{P1}^2 & \ldots & a_{PM}^2 \end{pmatrix} \begin{pmatrix} \varphi_1''\left(\sum_{p=1}^{P} a_{p1}t_p\right) \\ \vdots \\ \varphi_M''\left(\sum_{p=1}^{P} a_{pM}t_p\right) \end{pmatrix} = (A \cdot A) \begin{pmatrix} \varphi_1''\left(\sum_{p=1}^{P} a_{p1}t_p\right) \\ \vdots \\ \varphi_M''\left(\sum_{p=1}^{P} a_{pM}t_p\right) \end{pmatrix}$$

Assume Rank$(A \odot A) = M$.[28] Let $(A \odot A)^{+}$ be the Moore-Penrose pseudoinverse then

$$(A \odot A)^{+} \begin{pmatrix} \frac{\partial^2 \varphi_Y(t)}{\partial t_1^2} \\ \vdots \\ \frac{\partial^2 \varphi_Y(t)}{\partial t_{p_1} \partial t_{p_2}} \\ \vdots \\ \frac{\partial^2 \varphi_Y(t)}{\partial t_P^2} \end{pmatrix} = \begin{pmatrix} \varphi_1'' \left(\sum_{p=1}^{P} a_{p1} t_p \right) \\ \vdots \\ \varphi_M'' \left(\sum_{p=1}^{P} a_{pM} t_p \right) \end{pmatrix}$$

Denote the entries of $(A \odot A)^{+}$ by a_{pm}^{+}, $p = 1, \ldots, P^2$ and $m = 1, \ldots, M$. For any m^*

$$\sum_{p_1=1}^{P} \sum_{p_2=1}^{P} a_{p_1 p_2}^{+} \frac{\partial^2 \varphi_Y(t)}{\partial t_{p_1} \partial t_{p_1}} = \varphi_{m^*}''(\sum_{p=1}^{P} a_{pM} t_p)$$

For any (non-unique) t such that $u \sum_{p=1}^{P} a_{pM} t_p =: ue^{m^*} = u$ apply the second fundamental theorem of calculus twice

$$\varphi_{m^*}(s) = \int_0^s \int_0^v \sum_{k=1}^{p} \sum_{l=1}^{p} a_{p_1 p_2}^{+} \frac{\partial^2 \varphi_Y(ue^{m^*})}{\partial t_{p_1} \partial t_{p_2}} \mathrm{d}u \mathrm{d}v$$

Hence,

$$\phi_{m^*}(s) = \exp\left(\varphi_{m^*}(s)\right) = \exp\left(\int_0^s \int_0^v \sum_{k=1}^{p} \sum_{l=1}^{p} a_{p_1 p_2}^{+} \frac{\partial^2 \varphi_Y(ue^{m^*})}{\partial t_{p_1} \partial t_{p_2}} \mathrm{d}u \mathrm{d}v \right)$$

and

$$\frac{\partial^2 \varphi_Y(ue^{m^*})}{\partial t_{p_1} \partial t_{p_2}} = -\left(E\left[Y_{p_1} Y_{p_2} e^{iuY'e^{m^*}} \right] \phi_{Y'e^{m^*}}(u) + E\left[Y_{p_1} e^{iuY'e^{m^*}} \right] E\left[Y_{p_2} e^{iuY'e^{m^*}} \right] \right) \Big/ \left(\phi_{Y'e^{m^*}}(u)\right)^2$$

This identifies the characteristic function of U_{m^*} for $m^* = 1, \ldots, M$ in terms of the observed second order partial derivatives of log characteristic functions.

7.2 Proof of Lemma 1

Define the maximum norm $||Y||_\infty := \max\{|Y_1|, \ldots, |Y_P|\}$ and . Let $f_t(Y) = Y_p \exp(iY't)$, $p = 1, \ldots, P$

$$\Pr\left(\sup |E_N[f_t] - E[f_t]| > \varepsilon\right) = \Pr\left(\sup |E_N[f_t] - E[f_t]| > \varepsilon \,\Big|\, E_N ||Y||_\infty^2 \geq M \right) \cdot \Pr\left(E_N ||Y||_\infty^2 \geq M \right)$$

[28] This is the rank condition from Székely and Rao (2000) mentioned in the introduction that is necessary and sufficient for identification of U.

$$
\begin{aligned}
&+\Pr\left(\sup\left|E_N\left[f_t\right]-E\left[f_t\right]\right|>\varepsilon \,\middle|\, E_N\left\|Y\right\|_\infty^2<M\right)\cdot\Pr\left(E_N\left\|Y\right\|_\infty^2<M\right)\\
&\leq \Pr\left(E_N\left\|Y\right\|_\infty^2\geq M\right)+\Pr\left(\sup\left|E_N\left[f_t\right]-E\left[f_t\right]\right|>\varepsilon \,\middle|\, E_N\left\|Y\right\|_\infty^2<M\right)\\
&=A_1+A_2
\end{aligned}
$$

(i) Consider A_1

$$
\begin{aligned}
\Pr\left(E_N\left\|Y\right\|_\infty^2\geq M\right) &\leq E\left[\exp\left(E_N\left\|Y\right\|_\infty^2\right)\right]\Big/ e^M\\
&= E\left[\exp\left(\frac{1}{N}\sum_{n=1}^{N}\left\|Y_n\right\|_\infty^2\right)\right]\Big/ e^M\\
&= \prod_{n=1}^{N} E\left[\exp\left(\frac{1}{N}\left\|Y_n\right\|_\infty^2\right)\right]\Big/ e^M\\
&= E\left[e^{\frac{1}{N}\left\|Y_n\right\|_\infty^2}\right]^N\Big/ e^M\\
&= \left(1+\frac{E\left\|Y_n\right\|_\infty^2}{N}+\mathcal{O}\left(\frac{1}{N}\right)\right)^N\Big/ e^M
\end{aligned}
$$

where the first inequality follows from the use of a Chernoff bound, the second equality follows by independence, the third equality follows because the random variables are identically distributed and the last equality from a Taylor expansion around 0 as $\frac{1}{N}\to 0$ and the assumption that $\left\|Y_n\right\|_\infty^2$ is finite. Now notice that $\lim_{N\to\infty}\left[1+\frac{E\left\|Y\right\|_\infty^2}{N}+\mathcal{O}\left(\frac{1}{N}\right)\right]^N=e^{E\left\|Y\right\|_\infty^2}$ so choose N large enough so that

$$
\Pr\left(E_N\left\|Y\right\|_\infty^2\geq M\right)\leq\left(1+\frac{E\left\|Y\right\|_\infty^2}{N}+\mathcal{O}\left(\frac{1}{N}\right)\right)^N\Big/ e^M\leq\frac{2e^{E\left\|Y\right\|_\infty^2}}{e^M}
$$

(ii) To bound A_2 define the L_1-covering number, $N_1(\varepsilon,\mathcal{Q},\mathcal{F})$, as the smallest L for which there exist functions $g_1\ldots,g_L$ such that $\min_l E_{\mathcal{Q}}\left\|f-g_l\right\|\leq\varepsilon$ for all $f\in\mathcal{F}$ (Pollard (1984)).[29]

I will show that $N_1(\varepsilon,\mathcal{P_N},\mathcal{F})\leq C_3\left(\frac{TE_N\left\|Y_n\right\|_\infty^2}{\varepsilon}\right)^P$ where P_N is the empirical probability measure and $\mathcal{F}$ is the class of functions defined as $\mathcal{F}=\left\{f_t(Y):t\in[-T,T]^P\right\}$ where as before $f_t(Y)=Y_p\exp\left(iY't\right)$, $p=1,\ldots,P$. Discretize $[-T,T]^P$ into $L=\left(\frac{4TPE_N\left\|Y\right\|_\infty^2}{\varepsilon}\right)^P$ points, $t^1,\ldots,t^L$, by cutting $[-T,T]$ in each dimension into equidistant segments of length $\frac{\varepsilon}{2PE_N\left\|Y_n\right\|_\infty^2}$. Let $g_l(Y)=Y_p\exp\left(iY't^l\right)$ for $t^1,\ldots,t^L$ chosen above. For any $t\in[-T,T]^P$ there exists an l such that

$$
E_N\left|Y_p\exp\left(iY't\right)-Y_p\exp\left(iY't^l\right)\right|=E_N\left|Y_p\cos\left(Y't\right)+iY_p\sin\left(Y't\right)-Y_p\cos\left(Y't^l\right)-iY_p\sin\left(Y't^l\right)\right|
$$

[29] $\mathcal{Q}$ is a probability measure and $\mathcal{F}$ is a class of functions in $\mathcal{L}^1(\mathcal{Q})$

$$\begin{aligned}
&\leq E_N \left|Y_p \cos\left(Y't\right) - Y_p \cos\left(t'^l Y\right)\right| + E_N \left|iY_p \sin\left(Y't\right) - iY_p \sin\left(t'^l Y\right)\right| \\
&\leq 2PE_N \left|\left|Y_p \left(t'Y - t'^l Y\right)\right|\right|_\infty \\
&\leq 2PE_N \left|\left|Y_p Y\right|\right|_\infty \left|\left|t - t^l\right|\right|_\infty \\
&\leq 2PE_N \left|\left|Y\right|\right|_\infty^2 \left|\left|t - t^l\right|\right|_\infty \\
&\leq \varepsilon
\end{aligned}$$

It follows that the L_1-covering number satisfies $N_1(\varepsilon, P_N, \mathcal{F}) \leq C_3 \left(\frac{TE_N||Y||_\infty^2}{\varepsilon}\right)^P$ where $C_3 = (4P)^P$.

The L_1-covering ensures that the value of $f_t(Y) = Y_p \exp\left(iY't\right)$ for any point in $[-T, T]^P$ is arbitrarily close to at least one of $g_l(Y) = Y_p \exp\left(iY't^l\right)$. Next I use a result from Pollard (1984) who uses an exponential-type bound and the L_1-covering number to bound $\Pr\left(\sup\left|E\left[Y_p \exp\left(iY't\right)\right] - E_N\left[Y_p \exp\left(iY't\right)\right]\right|\right)$.

By assumption $E\left|\left|Y\right|\right|_\infty^2$ is bounded (and observed). Let $N \geq \frac{E||Y||_\infty^2}{8\varepsilon^2}$ then

$$\mathrm{Var}\left(E_N\left[f_t\right]\right) = \frac{1}{N}\mathrm{Var}\left(Y_p \exp\left(iY't^l\right)\right) \leq \frac{1}{N}E\left[Y_p^2\right] \leq \frac{1}{N}E\left|\left|Y\right|\right|_\infty^2 \leq 8\varepsilon^2$$

Equations (30) and (31) in Pollard (1984) now apply so that

$$\begin{aligned}
\Pr\left(\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| > \varepsilon \,\middle|\, E_N\left|\left|Y\right|\right|_\infty^2 < M\right) &\leq 8N_1\left(\varepsilon/8, P_N, \mathcal{F}\right) \exp\left(-\frac{N\varepsilon^2}{128} \Big/ \max_l E_N\left[g_l^2\right]\right) \\
&\leq C_4 \left(\frac{TE_N\left|\left|Y\right|\right|_\infty^2}{\varepsilon}\right)^P \exp\left(-\frac{N\varepsilon^2}{128} \Big/ E_N\left|\left|Y\right|\right|_\infty^2\right)
\end{aligned}$$

where $C_4 = 8C_3$ and the second inequality follows from

$$\max_l E_N\left[g_l^2\right] = \max_l\left[Y_p^2 \exp\left(i2Y't^l\right)\right] \leq E_N\left[Y_p^2\right] \leq E_N\left|\left|Y\right|\right|_\infty^2$$

Hence,

$$\Pr\left(\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| > \varepsilon \,\middle|\, E_N\left|\left|Y\right|\right|_\infty^2 < M\right) \leq C_4 \left(\frac{TM}{\varepsilon}\right)^P \exp\left(-\frac{N\varepsilon^2}{128M}\right)$$

For N large enough the bounds for A_1 and A_2 imply

$$\Pr\left(\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| > \varepsilon\right) \leq A_1 + A_2 \leq \frac{2e^{E||Y||_\infty^2}}{e^M} + C_4\left(\frac{TM}{\varepsilon}\right)^P \exp\left(-\frac{N\varepsilon^2}{128M}\right)$$

(iii) The last step is to apply the Borel-Cantelli Lemma so index ε, T and M by N and let

$$\begin{aligned} T_N &= C_1 N^{\delta/2} && C_1, \delta > 0 \\ \varepsilon_N &= C_2 \frac{\ln N}{\sqrt{N}} && \\ M_N &= (1+\alpha)\ln N && \alpha > 0 \end{aligned}$$

For $C_2^2 > 64(1+\alpha)(2+P(1+\delta))$ and N large enough

$$\begin{aligned} \Pr\left(\sup\left|E_N\left[f_t\right]-E\left[f_t\right]\right|>\varepsilon_N\right) &= \frac{2e^{E||Y||_\infty^2}}{e^{M_N}} + C_4\left(\frac{T_N M_N}{\varepsilon_N}\right)^P \exp\left(-\frac{N\varepsilon_N^2}{128 M_N}\right) \\ &= \frac{2e^{E||Y||_\infty^2}}{e^{M_N}} + C_4 \exp\left(P\ln\left(\frac{T_N M_N}{\varepsilon_N}\right) - \frac{N\varepsilon_N^2}{128 M_N}\right) \\ &= \frac{2e^{E||Y||_\infty^2}}{e^{(1+\alpha)\ln N}} + C_4 \exp\left(P\ln\left(\frac{C_1 N^{\delta/2}(1+\alpha)\ln N}{C_2\frac{\ln N}{\sqrt{N}}}\right) - \frac{N\left(C_2\frac{\ln N}{\sqrt{N}}\right)^2}{128(1+\alpha)\ln N}\right) \\ &= \frac{2e^{E||Y||_\infty^2}}{N^{1+\alpha}} + C_4 \exp\left(P\ln\left(\frac{C_1 N^{(\delta+1)/2}(1+\alpha)}{C_2}\right) - \frac{C_2^2\ln N}{128(1+\alpha)}\right) \\ &= \frac{2e^{E||Y||_\infty^2}}{N^{1+\alpha}} + C_4 \exp\left(P\ln\left(\frac{C_1(1+\alpha)}{C_2}\right) + \left[\frac{P\left(1+\delta\right)}{2} - \frac{C_2^2}{128(1+\alpha)}\right]\ln N\right) \\ &= \frac{2e^{E||Y||_\infty^2}}{N^{1+\alpha}} + C_5 \exp\left(\left[\frac{P\left(1+\delta\right)}{2} - \frac{C_2^2}{128(1+\alpha)}\right]\ln N\right) \\ &< \frac{2e^{E||Y||_\infty^2}}{N^{1+\alpha}} + C_5 \exp\left(-\ln N^{1+\beta}\right) \\ &\leq \frac{1}{N^{1+\min\{\alpha,\beta\}}}\left(2e^{E||Y||_\infty^2} + C_5\right) \\ &= \frac{C_6}{N^{1+\min\{\alpha,\beta\}}} \end{aligned}$$

where $C_5 = C_4\left(\frac{C_1(1+\alpha)}{C_2}\right)^P$, $C_6 = 2e^{E||Y||_\infty^2} + C_5$ and C_2^2 is chosen so that $\beta := -\frac{P(1+\delta)}{2} + \frac{C_2^2}{128(1+\alpha)} - 1 > 0$. The last equality follows from the assumptions that $E\left|\left|Y\right|\right|_\infty^2$ is bounded.

For the above choices of ε_N, T_N, M_N and C_2

$$\sum_{N=1}^{\infty} \Pr\left(\sup\left|E_N\left[f_t\right]-E\left[f_t\right]\right|>\varepsilon_N\right) < C_6 \sum_{N=1}^{\infty} \frac{1}{N^{1+\min\{\alpha,\beta\}}} < \infty$$

The Borel-Cantelli lemma then implies that

$$\sup\left|E_N\left[f_t\right]-E\left[f_t\right]\right| \leq \varepsilon_N \qquad \text{a.s}$$

for N large enough.

7.3 Proof of Lemma 2

I will show that $N_1(\varepsilon, \mathcal{P_N}, \mathcal{F}) \leq \widetilde{C}_3 \left(\frac{T}{\varepsilon}\right)^P$ where P_N is the empirical probability measure and $\mathcal{F}$ is the class of functions defined as $\mathcal{F} = \left\{\exp\left(iY't\right) : t \in [-T,T]^P\right\}$. Discretize $[-T,T]^P$ into $L = \left(\frac{4TP}{\varepsilon}\right)^P$ points, $t^1, \ldots, t^L$, by cutting $[-T,T]$ in each dimension into equidistant segments of length $\frac{\varepsilon}{2P}$. Let $g_l(Y) = \exp\left(iY't^l\right)$ for $t^1, \ldots, t^L$ chosen above. For any $t \in [-T,T]^P$ there exists an l such that

$$E_N \left|\exp\left(iY't\right) - \exp\left(iY't^l\right)\right| = E_N \left|\cos\left(Y't\right) + i\sin\left(Y't\right) - \cos\left(Y't^l\right) - i\sin\left(Y't^l\right)\right| \leq 2P\left|\left|t - t^l\right|\right|_\infty \leq \varepsilon$$

It follows that the L_1-covering number satisfies $N_1(\varepsilon, P_N, \mathcal{F}) \leq \widetilde{C}_3 \left(\frac{T}{\varepsilon}\right)^P$ where $\widetilde{C}_3 = (4P)^P$. By Pollard (1984)

$$\Pr\left(\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| > \varepsilon\right) \leq 8N_1\left(\varepsilon/8, P_N, \mathcal{F}\right)\exp\left(-\frac{N\varepsilon^2}{128}\Big/\max_l E_N\left[g_l^2\right]\right) \leq \widetilde{C}_4\left(\frac{T}{\varepsilon}\right)^P \exp\left(-\frac{N\varepsilon^2}{128}\right)$$

Index ε and T by N and let

$$T_N = \widetilde{C}_1\left(\frac{\ln N}{N}\right)^{\frac{1}{2}} N^{\frac{\widetilde{\delta}}{2}} \qquad \widetilde{C}_1 > 0, \widetilde{\delta} > 1$$

$$\varepsilon_N = \widetilde{C}_2\left(\frac{\ln N}{N}\right)^{\frac{1}{2}}$$

For $\widetilde{C}_2^2 > 64(2 + P\widetilde{\delta})$ and N large enough

$$\begin{aligned}
\Pr\left(\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| > \varepsilon_N\right) &= \widetilde{C}_4\left(\frac{T_N}{\varepsilon_N}\right)^P \exp\left(-\frac{N\varepsilon_N^2}{128}\right) \\
&= \widetilde{C}_4 \exp\left(P\ln\left(\frac{T_N}{\varepsilon_N}\right) - \frac{N\varepsilon_N^2}{128}\right) \\
&= \widetilde{C}_4 \exp\left(P\ln\left(\frac{\widetilde{C}_1\left(\frac{\ln N}{N}\right)^{\frac{1}{2}} N^{\frac{\widetilde{\delta}}{2}}}{\widetilde{C}_2\left(\frac{\ln N}{N}\right)^{\frac{1}{2}}}\right) - \frac{N\left(\widetilde{C}_2\left(\frac{\ln N}{N}\right)^{\frac{1}{2}}\right)^2}{128}\right) \\
&= \widetilde{C}_4 \exp\left(P\ln\left(\frac{\widetilde{C}_1}{\widetilde{C}_2}N^{\frac{\widetilde{\delta}}{2}}\right) - \frac{\widetilde{C}_2^2}{128}\ln N\right) \\
&= \widetilde{C}_4 \exp\left(P\ln\left(\frac{\widetilde{C}_1}{\widetilde{C}_2}\right) + \left[\frac{P\widetilde{\delta}}{2} - \frac{\widetilde{C}_2^2}{128}\right]\ln N\right) \\
&= \widetilde{C}_5 \exp\left(\left[\frac{P\widetilde{\delta}}{2} - \frac{\widetilde{C}_2^2}{128}\right]\ln N\right) \\
&< \widetilde{C}_5 \exp\left(-\ln N^{1+\widetilde{\beta}}\right) \\
&= \frac{\widetilde{C}_5}{N^{1+\widetilde{\beta}}}
\end{aligned}$$

where $\widetilde{C}_2^2$ is chosen so that $\widetilde{\beta} := -\frac{P\widetilde{\delta}}{2} + \frac{\widetilde{C}_2^2}{128} - 1 > 0$.

For the above choices of ε_N, T_N and $\widetilde{C}_2$

$$\sum_{N=1}^{\infty} \Pr\left(\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| > \varepsilon_N\right) < \widetilde{C}_5 \sum_{N=1}^{\infty} \frac{1}{N^{1+\widetilde{\beta}}} < \infty$$

The Borel-Cantelli lemma then implies that

$$\sup\left|E_N\left[f_t\right] - E\left[f_t\right]\right| \leq \varepsilon_N \qquad \text{a.s}$$

for N large enough.

7.4 Proof of Theorem 2

The proof requires four inequalities:

1. $\sup\left|\widehat{\phi}_{Yp}(t) - \phi_{Yp}(t)\right| \leq \varepsilon_N$ follows from Lemma 1.

2. $\sup\left|\widehat{\phi}_{Yp}(t)\right| = \sup\left|\widehat{\phi}_{Yp}(t) - \phi_{Yp}(t) + \phi_{Yp}(t)\right| \leq \sup|\left|\widehat{\phi}_{Yp}(t) - \phi_{Yp}(t)\right| + \sup|\phi_{Yp}(t)| \leq \varepsilon_N + E\,||Y||_{\infty}$ where the inequality follows from the triangle inequality and Lemma 1. By assumption $E\,||Y||_{\infty}$ is bounded.

3. For all $t \in [-T_N, T_N]^P$ $\left|\frac{\widehat{\phi}_Y(t) - \phi_Y(t)}{\phi_Y(t)}\right| \leq \frac{\varepsilon_N}{|\phi_Y(t)|}$ follows from Lemma 1.

4. For all $t \in [-T_N, T_N]^P$

$$\frac{1}{\left|1 + \frac{\widehat{\phi}_Y(t) - \phi_Y(t)}{\phi_Y(t)}\right|} \leq \frac{1}{\left|1 - \frac{\left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|}{|\phi_Y(t)|}\right|} = \frac{|\phi_Y(t)|}{\left||\phi_Y(t)| - \left|\widehat{\phi}_Y(t) - \phi_Y(t)\right|\right|} \leq \frac{|\phi_Y(t)|}{\left(|\phi_Y(t)| - \varepsilon_N\right)^{+}}$$

   where the first inequality follows from the triangle inequality and the second inequality follows from Lemma 1.

The proof proceeds as follows

$$\begin{aligned}
&\sup_{s\in[-S_N,S_N]}\left|\int_0^s \left(\frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\widehat{\phi}_Y(ue^{m^*})} - \frac{\phi_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})}\right)\mathrm{d}u\right| \\
&= \sup_{s\in[-S_N,S_N]}\left|\int_0^s \left(\frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\widehat{\phi}_Y(ue^{m^*})} - \frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} + \frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} - \frac{\phi_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})}\right)\mathrm{d}u\right| \\
&\leq \sup_{s\in[-S_N,S_N]}\left|\int_0^s \left(\frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\widehat{\phi}_Y(ue^{m^*})} - \frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})}\right)\mathrm{d}u\right| + \sup_{s\in[-S_N,S_N]}\left|\int_0^s \left(\frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} - \frac{\phi_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})}\right)\mathrm{d}u\right|
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{s\in[-S_N,S_N]} \left| \int_0^s - \left(\frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} \cdot \frac{\frac{\widehat{\phi}_Y(ue^{m^*})-\phi_Y(ue^{m^*})}{\phi_Y(ue^{m^*})}}{1+\frac{\widehat{\phi}_Y(ue^{m^*})-\phi_Y(ue^{m^*})}{\phi_Y(ue^{m^*})}} \right) \mathrm{d}u \right| + \sup_{s\in[-S_N,S_N]} \left| \int_0^s \left(\frac{1}{\phi_Y(ue^{m^*})} \left(\widehat{\phi}_{Yp}(ue^{m^*}) - \phi_{Yp}(ue^{m^*}) \right) \right) \mathrm{d}u \right| \\
&\leq \sup_{s\in[-S_N,S_N]} \int_0^s \left| \widehat{\phi}_{Yp}(ue^{m^*}) \right| \left| \frac{1}{\phi_Y(ue^{m^*})} \right| \left| \frac{\frac{\widehat{\phi}_Y(ue^{m^*})-\phi_Y(ue^{m^*})}{\phi_Y(ue^{m^*})}}{1+\frac{\widehat{\phi}_Y(ue^{m^*})-\phi_Y(ue^{m^*})}{\phi_Y(ue^{m^*})}} \right| \mathrm{d}u + \sup_{s\in[-S_N,S_N]} \int_0^s \left| \frac{1}{\phi_Y(ue^{m^*})} \right| \left| \widehat{\phi}_{Yp}(ue^{m^*}) - \phi_{Yp}(ue^{m^*}) \right| \mathrm{d}u \\
&\leq \sup \left| \widehat{\phi}_{Yp}(ue^{m^*}) \right| \sup_{s\in[-S_N,S_N]} \int_0^s \left| \frac{1}{\phi_Y(ue^{m^*})} \right| \frac{\sup \frac{\left|\widehat{\phi}_Y(ue^{m^*})-\phi_Y(ue^{m^*})\right|}{\left|\phi_Y(ue^{m^*})\right|}}{\left| 1+\frac{\widehat{\phi}_Y(ue^{m^*})-\phi_Y(ue^{m^*})}{\phi_Y(ue^{m^*})} \right|} \mathrm{d}u + \sup \left| \widehat{\phi}_{Yp}(ue^{m^*}) - \phi_{Yp}(ue^{m^*}) \right| \sup_{s\in[-S_N,S_N]} \int_0^s \left| \frac{1}{\phi_Y(ue^{m^*})} \right. \\
&\leq \left(\varepsilon_N + E\,||Y||_\infty \right) \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})|} \cdot \frac{\varepsilon_N}{|\phi_Y(ue^{m^*})|} \cdot \frac{\left|\phi_Y(ue^{m^*})\right|}{(|\phi_Y(ue^{m^*})| - \varepsilon_N)^+} \mathrm{d}u + \varepsilon_N \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u \\
&= \varepsilon_N \left(\varepsilon_N + E\,||Y||_\infty \right) \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})| \left(|\phi_Y(ue^{m^*})| - \varepsilon_N \right)^+} \mathrm{d}u + \varepsilon_N \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u \\
&= \varepsilon_N \left(\varepsilon_N + E\,||Y||_\infty \right) \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})| \left(|\phi_Y(ue^{m^*})| - \varepsilon_N \right)^+} \mathrm{d}u + \varepsilon_N \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})|} \mathrm{d}u \\
&\leq C_7 \varepsilon_N \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})| \left(|\phi_Y(ue^{m^*})| - \varepsilon_N \right)^+} \mathrm{d}u
\end{aligned}
$$

where the fourth inequality follows from the four inequalities stated before the proof.

For large N, $\sup_{s\in[-S_N,S_N]} \left| \int_0^s \frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\widehat{\phi}_Y(ue^{m^*})} \mathrm{d}u - \int_0^s \frac{\phi_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} \mathrm{d}u \right| \leq 1$ so that

$$
\begin{aligned}
\sup_{s\in[-S_N,S_N]} \left| \widehat{\phi}_{m^*}(s) - \phi_{m^*}(s) \right| &= \sup_{s\in[-S_N,S_N]} \left| \exp\left(\int_0^s \frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\widehat{\phi}_Y(ue^{m^*})} \mathrm{d}u \right) - \exp\left(\int_0^s \frac{\phi_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} \right) \mathrm{d}u \right| \\
&\leq \sup_{s\in[-S_N,S_N]} \left| \int_0^s \frac{\widehat{\phi}_{Yp}(ue^{m^*})}{\widehat{\phi}_Y(ue^{m^*})} \mathrm{d}u - \int_0^s \frac{\phi_{Yp}(ue^{m^*})}{\phi_Y(ue^{m^*})} \mathrm{d}u \right| \\
&\leq C_7 \varepsilon_N \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ue^{m^*})| \left(|\phi_Y(ue^{m^*})| - \varepsilon_N \right)^+} \mathrm{d}u
\end{aligned}
$$

7.5 Proof of Theorem 3

For all u in the support of U_{m^*} and for N large enough

$$
\begin{aligned}
&\left| \widehat{f}_{m^*}(u) - f_{m^*}(u) \right| \\
&= \left| \frac{1}{2\pi} \int e^{-isu} \widehat{\phi}_{m^*}(s) \phi_K(sh_N) \, \mathrm{d}s - \frac{1}{2\pi} \int e^{-isu} \phi_{m^*}(s) \mathrm{d}s \right| \\
&= \left| \frac{1}{2\pi} \int e^{-isu} \left(\widehat{\phi}_{m^*}(s) \phi_K(sh_N) - \phi_{m^*}(s) \phi_K(sh_N) + \phi_{m^*}(s) \phi_K(sh_N) - \phi_{m^*}(s) \right) \mathrm{d}s \right| \\
&= \left| \frac{1}{2\pi} \int e^{-isu} \phi_K(sh_N) \left(\widehat{\phi}_{m^*}(s) - \phi_{m^*}(s) \right) \mathrm{d}s + \frac{1}{2\pi} \int e^{-isu} \phi_{m^*}(s) \left(\phi_K(sh_N) - 1 \right) \mathrm{d}s \right| \\
&\leq \frac{1}{2\pi} \int |\phi_K(sh_N)| \left| \widehat{\phi}_{m^*}(s) - \phi_{m^*}(s) \right| + \frac{1}{2\pi} \int |\phi_{m^*}(s)| \, |\phi_K(sh_N) - 1| \, \mathrm{d}s \\
&\leq \frac{1}{2\pi} \int_{-S_N}^{S_N} \left| \widehat{\phi}_{m^*}(s) - \phi_{m^*}(s) \right| \mathrm{d}s + \frac{1}{2\pi} \int_{-S_N}^{S_N} |\phi_{m^*}(s)| \, |m(sh_N) \, (sh_N)^q| \, \mathrm{d}s + \frac{1}{\pi} \int_{S_N}^{\infty} |\phi_{m^*}(s)| \, \mathrm{d}s \\
&\leq \frac{S_N}{\pi} \sup_{s\in[-S_N,S_N]} \left| \widehat{\phi}_{m^*}(s) - \phi_{m^*}(s) \right| + \frac{1}{2\pi} \sup_{t\in[-1,1]} |m(s)| \, h_N^q \int_{-S_N}^{S_N} |\phi_{m^*}(s)| |s|^q \mathrm{d}s + \frac{1}{2\pi} \int_{-\infty}^{-S_N} |\phi_{m^*}(s)| \mathrm{d}s + \frac{1}{2\pi} \int_{S_N}^{\infty} |\phi_{m^*}(s)| \mathrm{d}s
\end{aligned}
$$

$$\leq \frac{S_N C_7 \varepsilon_N}{\pi} \int_{-S_N}^{S_N} \frac{1}{|\phi_Y(ve^*)| \left(|\phi_Y(ve^*)| - \varepsilon_N\right)^+} \mathrm{d}v + \frac{1}{2\pi} \sup_{s \in [-1,1]} |m(s)| \, h_N^q \int_{-S_N}^{S_N} |s|^q |\phi_{m^*}(s)| \mathrm{d}s + \frac{1}{2\pi} \int_{-\infty}^{-S_N} |\phi_{m^*}(s)| \mathrm{d}s + \frac{1}{2\pi} \int_{S_N}^{\infty} |\phi_{m^*}($$

where the second inequality follows because $|\phi_K(s)| < 1$, $\phi_K(s) = 1 + m(s)s^q$ for $s \in [-1, 1]$ and $\phi_K(s) = 0$ otherwise and $m(s)$ is continuous for $s \in [-1, 1]$, the third inequality follows because m is continuous on a compact interval and the fourth inequality follows from Theorem 1.

REFERENCES

BONDESSON, L. (1974). "Characterizations of Probability Laws Through Constant Regression," *Z. Wahrsch. v. Geb*, **29**, 93-115.

BONHOMME, S., and ROBIN, J.-M. (2010), "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *Review of Economic Studies*, **77** (2), 491-533.

CARRASCO, M., FLORENS J.-P. (2010), "Spectral Method for Deconvolving a Density", *Econometric Theory*, forthcoming

CARROLL, R.J., RUPPERT, D., STEFANSKI, L.A., CRAINICEANU, C. (2006) , *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition* (Chapman and Hall).

CHEN, X., HONG, NEKIPELOV, D. (2011), "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, forthcoming

CHERNOZHUKOV, V., FERNANDEZ-VAL, I., HAHN, J. and NEWEY, W. (2010), "Average and Quantile Effects in Nonseparable Panel Models"

CHESHER, A. (2007), "Instrumental Values," *Journal of Econometrics*, **139**, 15-34.

CSÖRGO, S. (1981), "Limit Behaviour of the Empirical Characteristic Function," *Annals of Probability*, **9** (1), 130-144.

CUNHA, F., HECKMAN, J., SCHENNACH, S. M., (2010), "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*. **78**, 883-931.

DELAIGLE, A. and GIJBELS, I. (2002), "Estimation of Integrated Squared Density Derivatives from a Contaminated Sample," *Journal of the Royal Statistical Society, Series B*, **64**, 869886.

DELAIGLE, A., HALL, P. and MEISTER, A. (2008), "On Deconvolution with Repeated Measurements," *Annals of Statistics*, **36** (2), 665685.

EVDOKIMOV, K. (2011), "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity"

EVDOKIMOV, K. White, H. (2011). "An Extension of a Lemma of Kotlarski,"

FAN, J. Q. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Annals of Statistics*, **19**, 1257-1272.

GAUTIER, E. and KITAMURA, Y. (2009) "Nonparametric Estimation in Random Coefficients Binary Choice Models,"

HOROWITZ, J. L. and MARKATOU, M. (1996), "Semiparametric Estimation of Regression Models for Panel Data," Review of Economic Studies, 63, 145168.

HSIAO, C. (1986). *Analysis of Panel Data* (Cambridge: Cambridge University Press).

HU, Y. and RIDDER, G. (2011), "Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information," *Journal of Applied Econometrics*, forthcoming

HU, Y. and RIDDER, G. (2010), "On Deconvolution as a First Stage Nonparametric Estimator," *Econometric Reviews*, **29**, 1-32.

HU, Y. and SCHENNACH, S. M. (2007), "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, **75**, 201-239.

KHATRI, C. G. and RAO, C. R. (1968), "Solutions to Some Functional Equations and their Applications to Characterization of Probability Distributions," *Sankhyä*, **30**, 167180.

KOTLARSKI, I. (1967), "On Characterizing the Gamma and Normal Distribution," *Pacific Journal of Mathematics*, **20**, 6976.

LEVINSOHN, J. and PETRIN, A.(2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 317-342.

LI, T., PERRIGNE, I. and VUONG, Q. (2000), "Conditionally Independent Private Information in OSC Wildcat Auctions," *Journal of Econometrics*, **98**, 129-161.

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

MATZKIN, R. L. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, **71** (5), 1339-1375.

MEISTER, A. (2007) "Deconvolving Compactly Supported Densities." *Mathematical Methods of Statistics*, **16**, 63-76.

OLLEY, S. and PAKES, A. (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, **64**, 1263-1295.

POLLARD, D. (1984), *Convergence of Stochastic Processes* (New York: Springer).

RAO, C. R. (1971) "Characterization of Probability Laws by Linear Functions," *Sankhyä*, **33**, 265-270.

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

SZÉKELY, G. J. and RAO, C. R. (2000), "Identifiability of Distributions of Independent Random Variables by Linear Combinations and Moments," *Sankhyä*, **62**, 193-202.

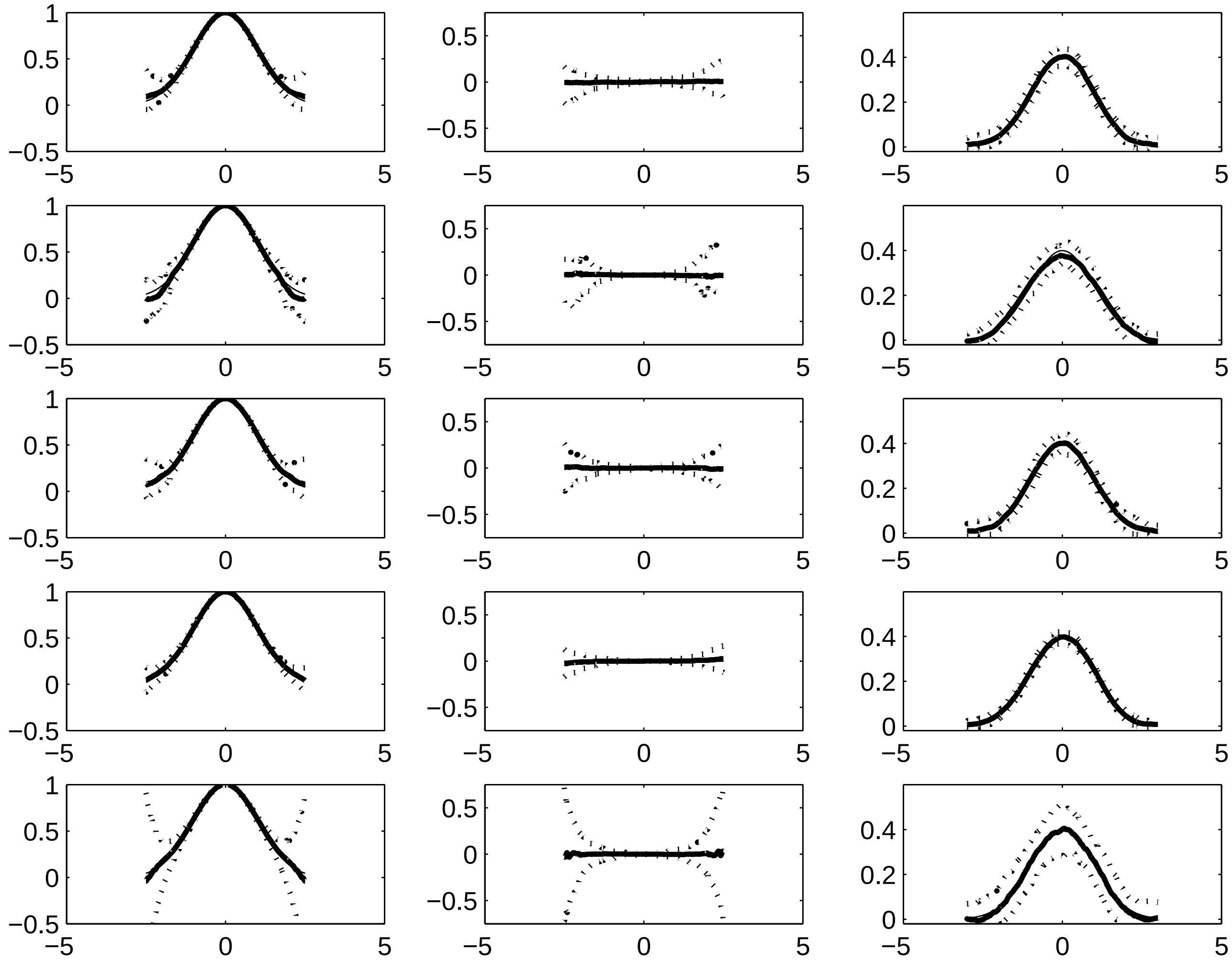

Figure 1: Measurement error model with repeated measurements. Experiment 1: $X^* \sim \text{Normal}(0, 1)$ with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

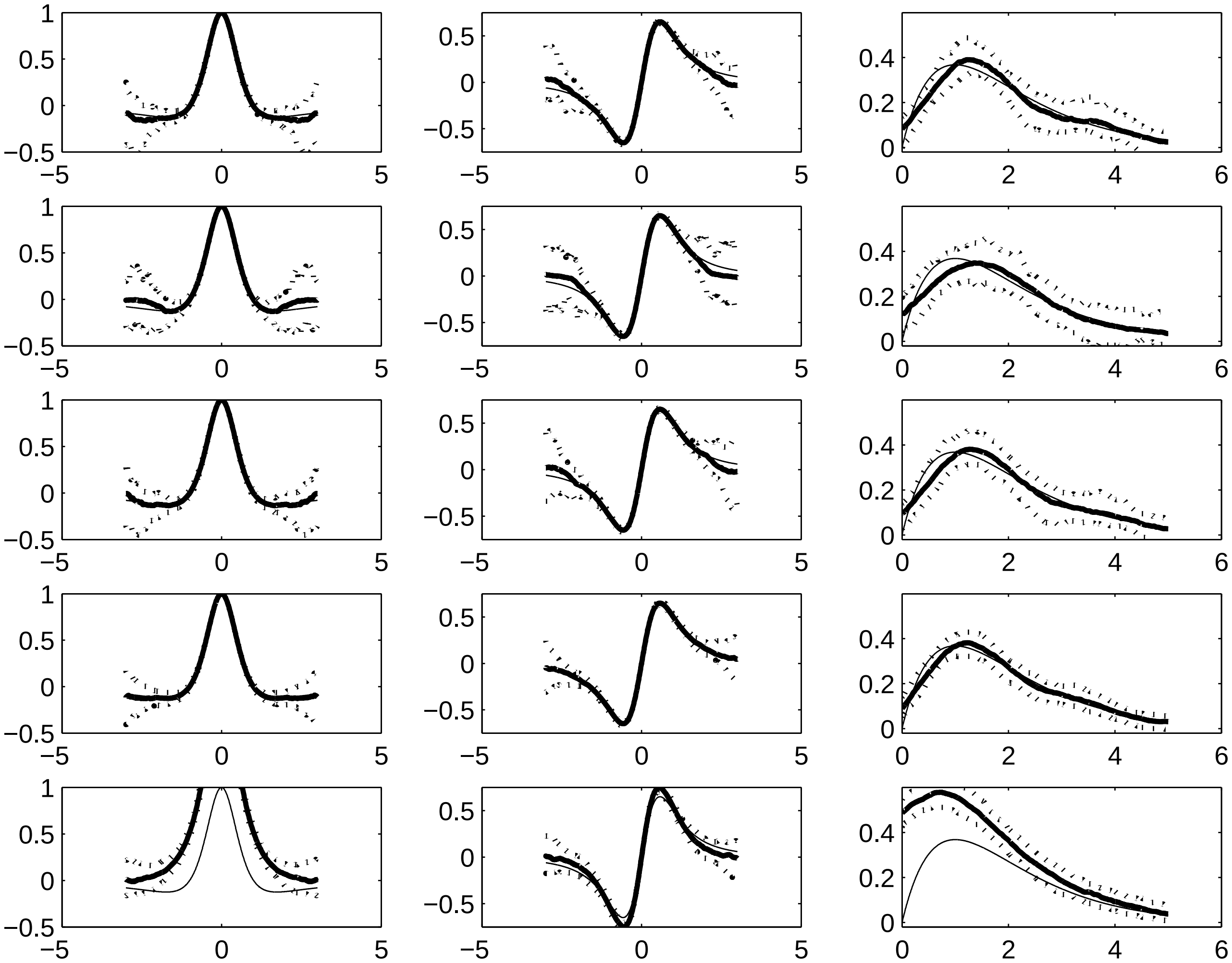

Figure 2: Measurement error model with repeated measurements. Experiment 2: $X^* \sim \text{Gamma}(2, 1)$ with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

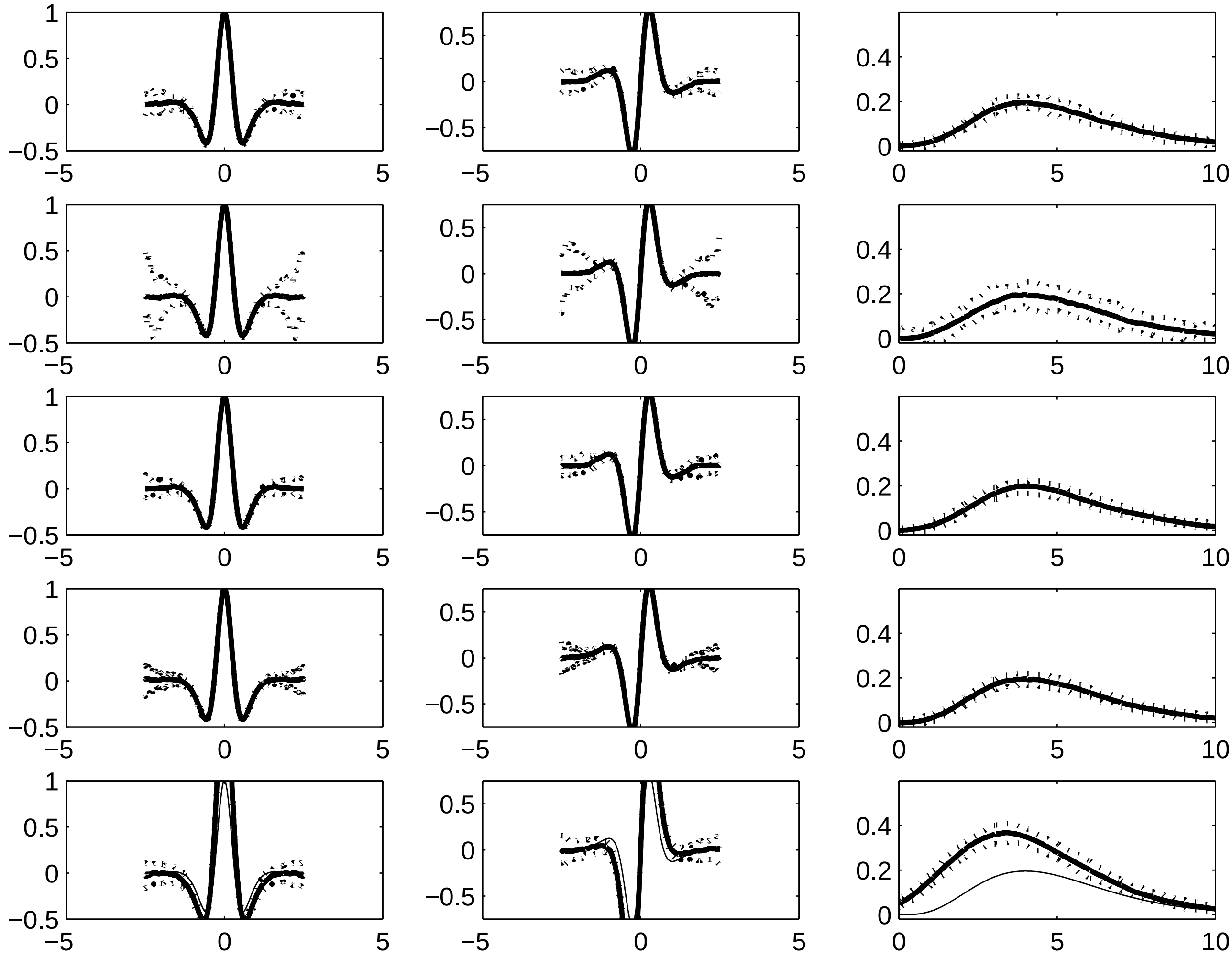

Figure 3: Measurement error model with repeated measurements. Experiment 3: $X^* \sim \text{Gamma}(5, 1)$ with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

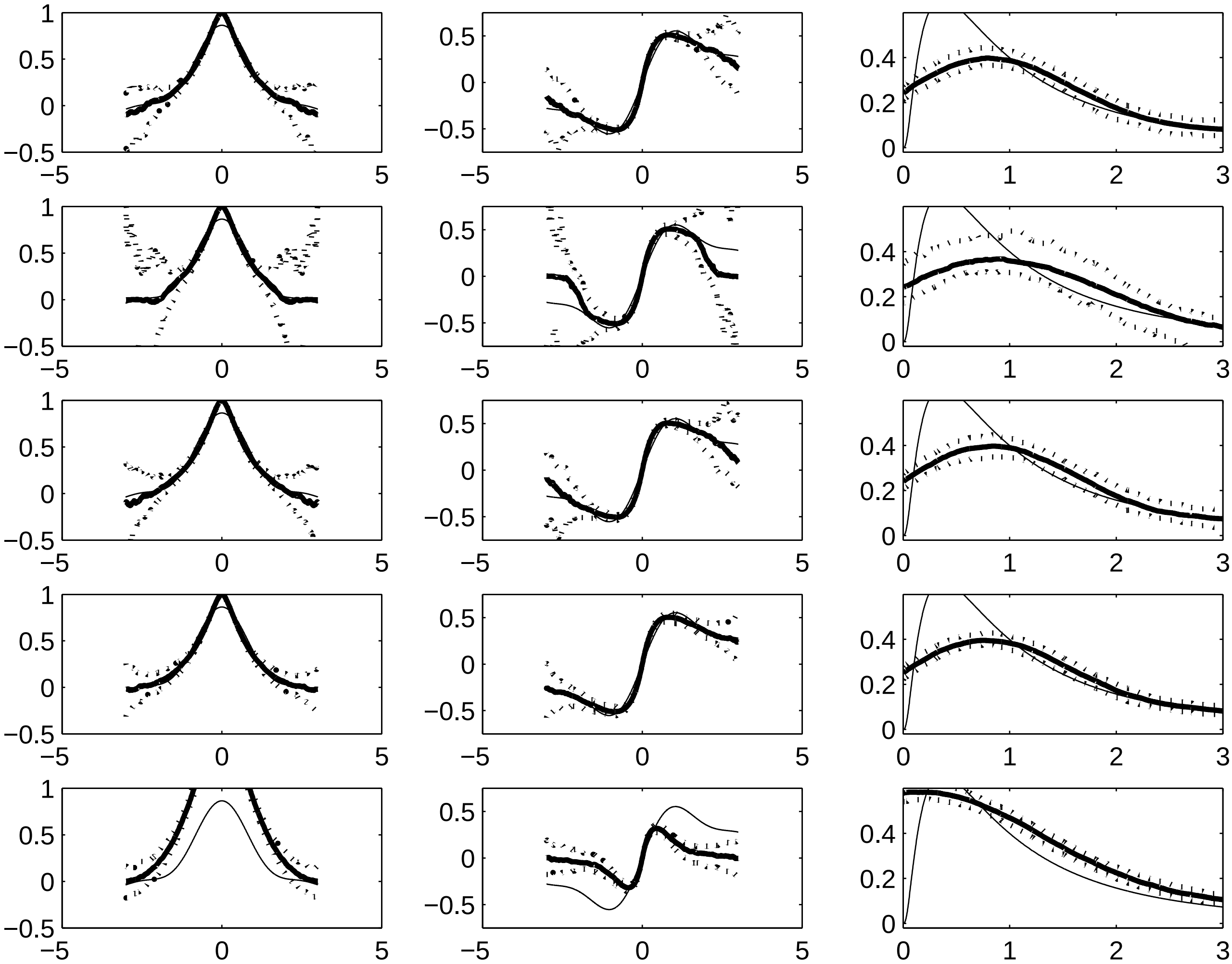

Figure 4: Measurement error model with repeated measurements. Experiment 4: X^* ∼ Lognormal with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

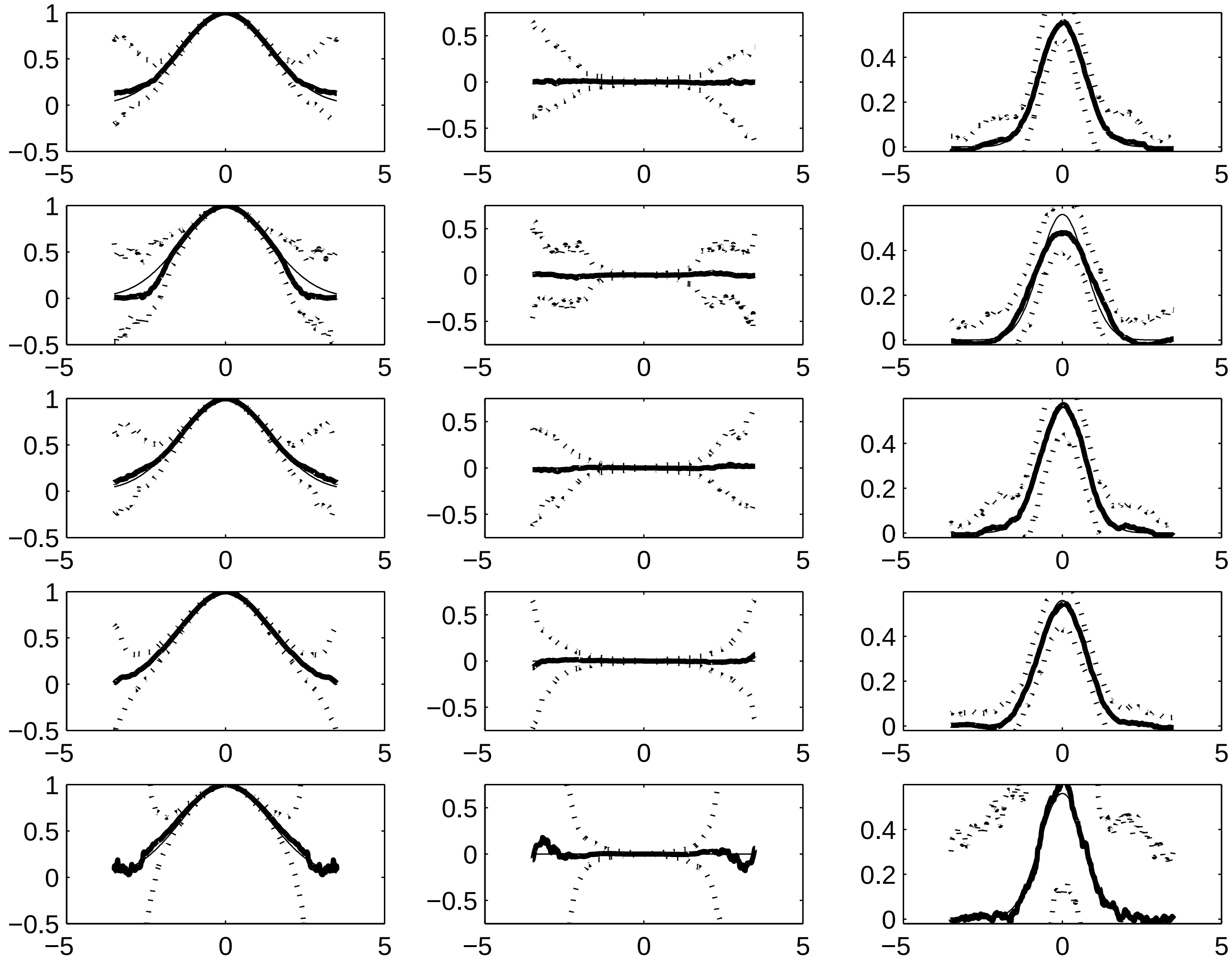

Figure 5: Measurement error model with repeated measurements. Experiment 5: $X^* \sim \frac{400}{403}N(0, \frac{1}{2}) + \frac{3}{403}N(0, \frac{406}{6})$ (Unimodal with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

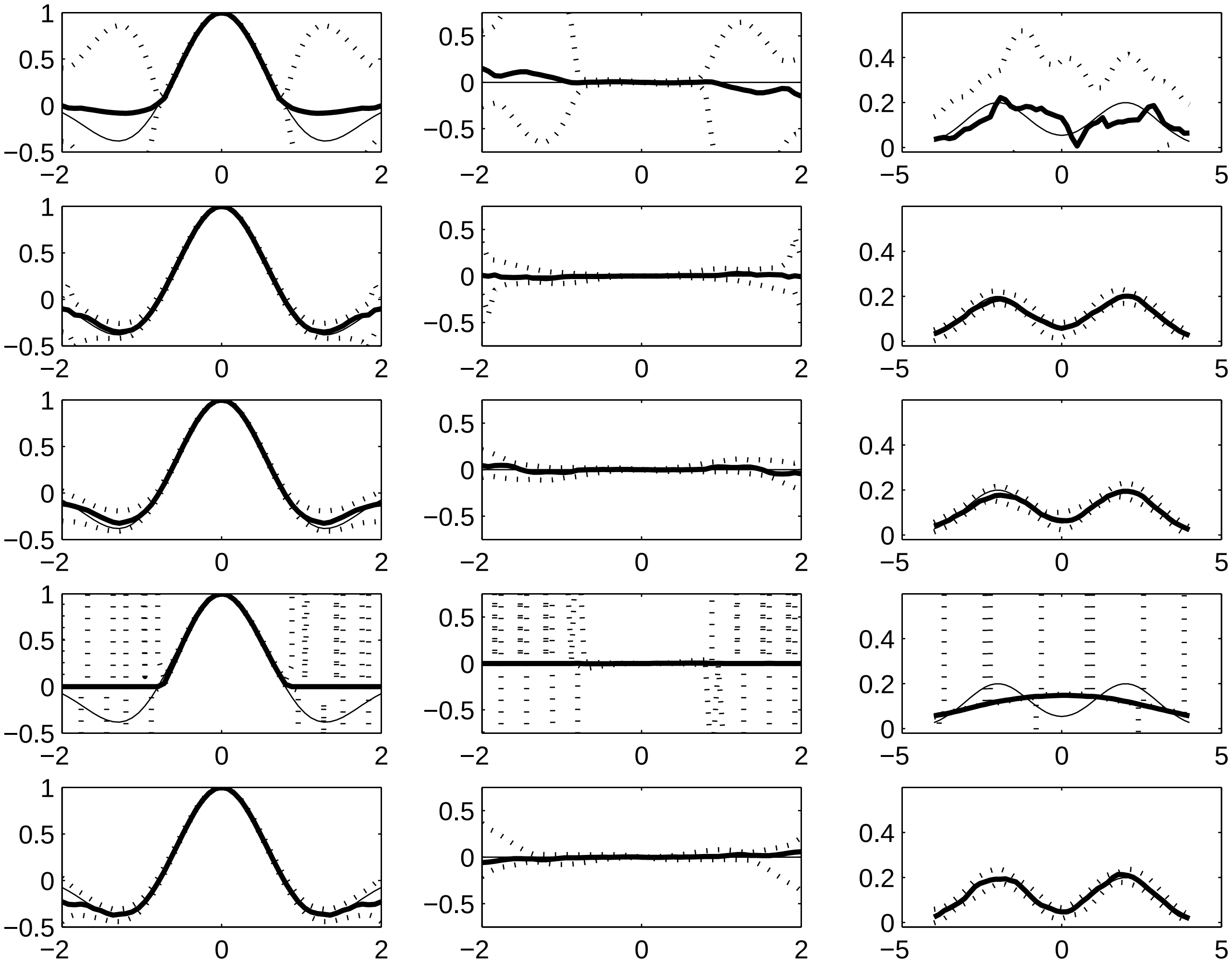

Figure 6: Measurement error model with repeated measurements. Experiment 6: $X^* \sim \frac{1}{2}N(-2,1) + \frac{1}{2}N(2,1)$ (Bimodal) with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

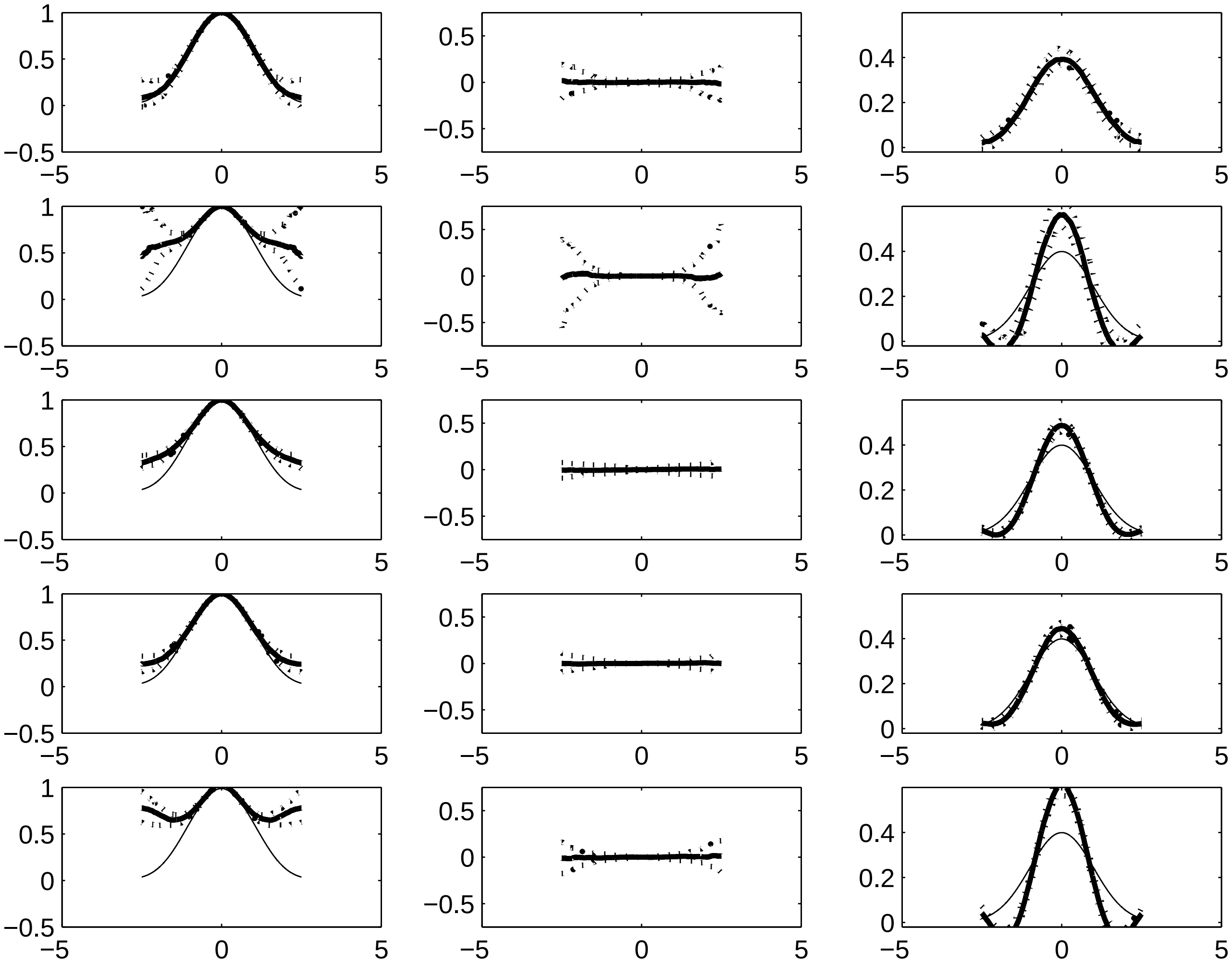

Figure 7: Measurement error model with repeated measurements. Experiment 7: $X^* \sim \mathrm{Normal}(0,1)$ (X^* and ε_1 dependent) with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

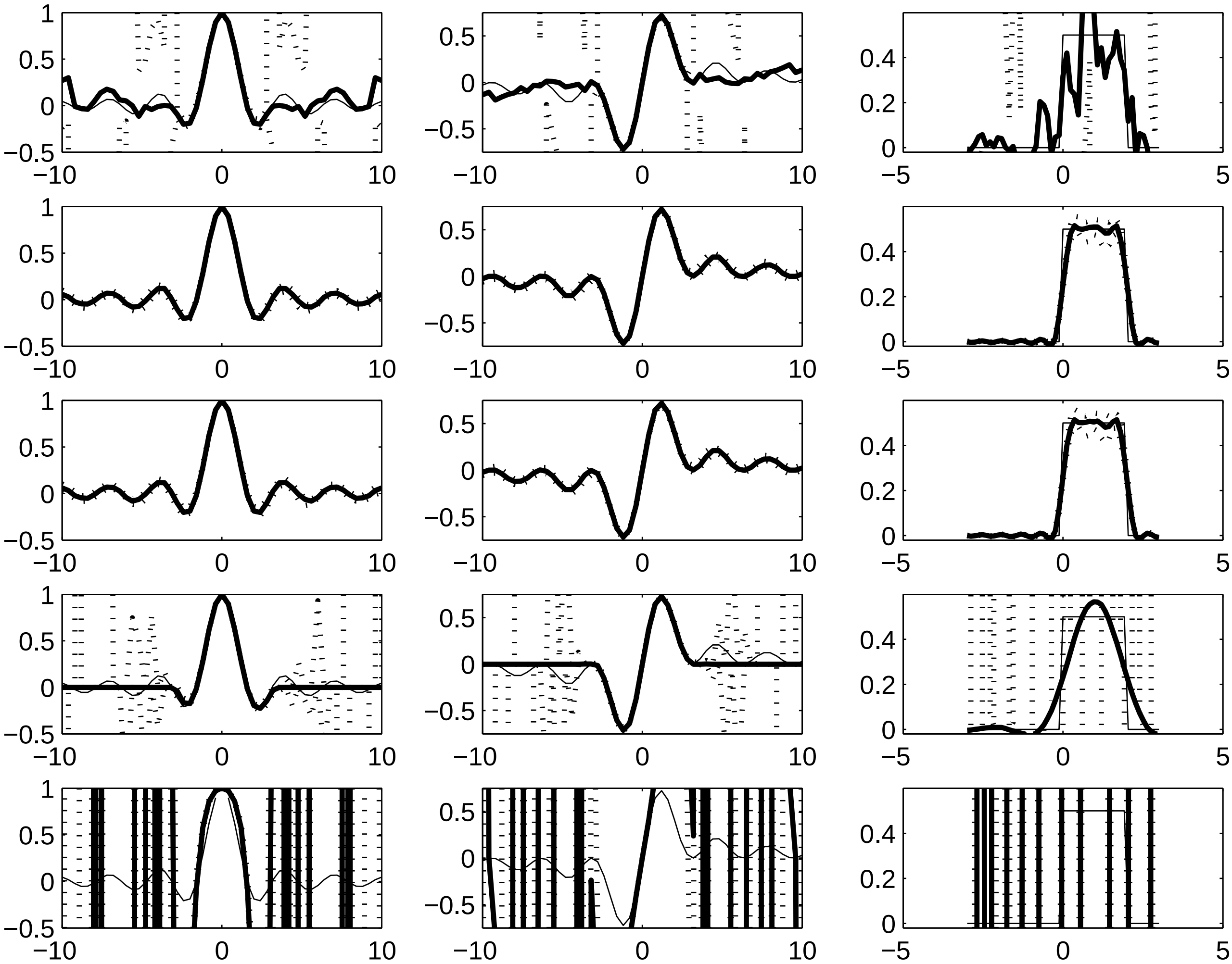

Figure 8: Measurement error model with repeated measurements. Experiment 8: $X^* \sim \text{Unif}(0,2)$ with $N = 1,000$. The left column is the real part of the characteristic function, the middle column is the imaginary part of the characteristic function and the right column is the density. The first through fifth rows are estimators 1 though 5, respectively.

Table 1: Comparing Estimators (N=100)

Experiment		Estimator A	Estimator B	Estimator C	Estimator D	Estimator E
Norm(0,1)	MISE	3.4838	0.0653	0.0707	0.0254	0.6088
Gamma(2,1)	MISE	82.853	2.6444	0.8068	4424.4	0.2001
Gamma(5,1)	MISE	49.543	0.5971	0.3186	NaN	0.2088
Lognormal	MISE	43.808	1.9967	0.0962	944.75	0.0612
Unimodal	MISE	1148.0	700.19	5.1063	>1,000	4.2132
Bimodal	MISE	0.0351	0.2117	0.0238	NaN	0.0786
Norm(0,1) (Depend)	MISE	0.0284	0.2545	0.0294	0.0178	0.1585
U(0,1)	MISE	>1,000	0.0329	0.0323	NaN	430.29

Table 2: Comparing Estimators (N=1,000)

Experiment		Estimator A	Estimator B	Estimator C	Estimator D	Estimator E
N(0,1)	MISE	0.0059	0.0128	0.0067	0.0022	0.0319
Gamma(2,1)	MISE	53.144	1.1908	0.0234	0.0081	0.1227
Gamma(5,1)	MISE	0.1781	0.0957	139.63	>1,000	0.1417
Lognormal	MISE	0.0467	18.107	0.0487	0.0422	0.0396
Unimodal	MISE	3.2454	743.32	0.4366	0.0320	1.1128
Bimodal	MISE	1.8272	0.0068	0.0026	NaN	0.0037
Norm(0,1) (Depend)	MISE	0.0036	0.0771	0.0158	0.0056	0.0799
U(0,1)	MISE	0.2322	0.0174	0.0174	NaN	256.60

Table 3: Comparing Estimators (N=10,000)

Experiment		Estimator A	Estimator B	Estimator C	Estimator D	Estimator E
N(0,1)	MISE	0.0007	0.0021	0.0006	0.0003	0.0023
Gamma(2,1)	MISE	1.0548	0.0090	0.0106	0.0023	0.1131
Gamma(5,1)	MISE	0.0010	0.0132	0.0007	0.0003	0.1359
Lognormal	MISE	0.0975	0.0792	8.4942	0.0408	0.0394
Unimodal	MISE	0.0409	0.6248	0.0267	>1,000	0.2490
Bimodal	MISE	54.579	0.0006	0.0004	NaN	0.0004
Norm(0,1) (Depend)	MISE	0.0005	0.0671	0.0149	0.0044	0.0692
U(0,1)	MISE	>1,000	0.0158	0.0158	NaN	238.12