Recent Developments in Semi-/Non-Parametric Estimation of Censoring, Sample Selection, Missing Data, and Measurement Error in Panel Data

Yvette Yu Zhang*, Qi Li†, Dong Li‡

May 10, 2011

Abstract

This article reviews the recent developments in the estimation of panel data models, in which some variables are only partially observed. Specially we consider the issues of censoring, sample selection, attribution, missing data and measurement errors in panel data models. Although most of these issues, except attribution, occur in cross sectional or time series data as well, panel data models introduce some particular challenges due to the presence of persistent individual effects. The past two decades have seen many stimulating developments in the econometric and statistical methods in dealing with these problems. This review focuses on two strands of research of the rapidly growing literature on semi- or non-parametric methods for panel data models: (i) estimations of panel models with discrete/limited dependent variables, and (ii) estimations of panel models based on nonparametric deconvolution methods.

*Department of Agricultural Economics, Texas A&M University; email: yzhang@tamu.edu
†Department of Economics, Texas A&M University; email: qi@econmail.tamu.edu
‡Department of Economics, Kansas State University; email: dongli@ksu.edu
1 Introduction

This article reviews the recent developments in the estimation of panel data models in which some variables are only partially observed. Some of the issues, such as censoring, sample selection, missing data and measure errors in variables, manifest themselves in other sampling schemes, such as cross sectional or time series data as well. On the other hand, attrition is unique in panel data. The complications associated with the incomplete information, combined with the unique structure of panel data models, can be rather challenging and calls for particular care in model construction and estimation.

Censoring has been extensively studied by economics. Tobin (1958) considered the relationship between expenditures on durable goods and disposable income. Even if the underlying latent variable enters the model linearly, the observed expenditures enter the model nonlinearly because we only observe zero or positive expenditures on durable goods. Tobin (1958) proposed a maximum likelihood estimator for the model, which was latter dubbed the Tobit model by Amemiya (1985). Amemiya (1985) further coined the terms Type I to Type V Tobit models that gained some popularity in econometrics literature. According to his terminology, Heckman (1976)’s two-step estimator is in fact a Type III Tobit model. Other early studies on the topic include Heckman (1979), Nelson (1977) Maddala (1980, 1983), Hausman and Wise (1979), Ridder (1990, 1992). As panel data models become increasingly popular in economic research, economists carry over the research on censoring and sample selection from cross-sectional data to panel data. With more and more micro panel data sets are available, attrition becomes a prominent problem. Cameron and Trivedi (2005, pp. 800-801) briefly reviewed the parametric methods on these models. Wooldridge (2010, Chapter 19) provided an in-depth coverage of the topics in both parametric and non/semi-parametric estimations. Verbeek and Nijman (1996) presented an early review of these models in panel data. More recently, Honoré, Vella, and Verbeek (2008) provided a comprehensive review on
both parametric and non/semi-parametric methods in panel data.

A second focus of this review is deconvolution-based methods in panel data models. Deconvolution is commonly used in estimations with measurement errors in variables. This method lends itself naturally to panel data models. For example, the composite error terms in panel data models can be viewed as the convolution between a persistent individual effect $\mu_i$ and an idiosyncratic error $\nu_{it}$. This is the approach taken by Horowitz and Markatou (1996). Alternatively, Evdokimov (2010) treats the observed outcome $Y_{it}$ in a panel data as the convolution between a conditional mean $m(x_{it}, \mu_i)$ and the error $\nu_{it}$, where the structural part is modeled nonparametrically without assuming the separability between the conditional mean $E[Y_{it}|x_{it}]$ and $\mu_i$. We go through these two papers in details to familiarize readers with this line of research.

2 Models with incomplete data

In this section we review estimators for panel data methods with censoring, sample selection or missing data issues. This section draws heavily on Li and Racine (2007) and Ai and Li (2008). In estimating linear panel data models, in which the dependent variable is a linear function of the unobservable individual effect, we usually eliminate the individual effects through first-differencing or the within transformation. But in all the models considered in this section, the observed dependent variable is a nonlinear function of the individual effect and thus no simple transformation is available to remove the individual effects.
2.1 Censoring

Consider the following panel data censored regression model:

\[ y_{it}^* = x_{it}' \beta + \mu_i + \nu_{it}, \quad (1) \]
\[ y_{it} = \max\{0, y_{it}^*\}, \quad i = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, T. \quad (2) \]

The observed dependent variable, \( y_{it} \), is nonlinear in the unobservable individual effects. So the usual techniques in removing the unobservable individual effects (\( \mu_i \)'s), including the within transformation (\( y_{it} - \bar{y}_i \)) and the time difference transformation (\( y_{it} - y_{is} \)) cannot remove the unobservable individual effects.

Applying a simple time-difference yields

\[ (y_{it} - x_{it}' \beta) - (y_{is} - x_{is}' \beta) = \max\{\mu_i + \nu_{it}, -x_{it}' \beta\} - \max\{\mu_i + \nu_{is}, -x_{is}' \beta\}. \quad (3) \]

If the right-hand-side of the above equation has a zero conditional mean, standard regression techniques can estimate the parameters consistently. For the right-hand-side to have a zero conditional mean, the two sufficient conditions are: (i) conditional on \((x_{it}, x_{is}, \mu_i), \nu_{it}\) and \(\nu_{is}\) are identically distributed; and (ii) the censoring values \(-x_{it}' \beta\) and \(-x_{is}' \beta\) are identical.

Condition (i) can be imposed as an assumption on the model. Condition (ii) is usually satisfied through a nonlinear differencing by artificially censoring the observed dependent variable so that the two error terms are censored at the same value of \(\max\{-x_{it}' \beta, -x_{is}' \beta\}\).
For that purpose, define

\[ e(y_{it} - x_{it}'\beta, x_{is}'\beta) = \max\{y_{it} - x_{it}'\beta, -x_{is}'\beta\} \]
\[ = \max\{\mu_i + \nu_{it}, -x_{it}'\beta, -x_{is}'\beta\}, \]
\[ e(y_{is} - x_{is}'\beta, x_{it}'\beta) = \max\{y_{is} - x_{is}'\beta, -x_{it}'\beta\} \]
\[ = \max\{\mu_i + \nu_{is}, -x_{is}'\beta, -x_{it}'\beta\}. \]

Then the conditional expectation of the difference between the above two terms is zero:

\[ \mathbb{E}\{e(y_{it} - x_{it}'\beta, x_{is}'\beta) - e(y_{is} - x_{is}'\beta, x_{it}'\beta)|x_{it}, x_{is}\} = 0. \] (4)

A GMM estimator can be constructed based on the above moment condition. But moment conditions based on zero conditional mean are usually more complicated than those based on zero unconditional mean. A better alternative is to find a convex objective function \( \rho \) so that its first order condition coincides with some unconditional moment conditions implied by (4). For example, consider the following function

\[ A(\beta) \equiv \mathbb{E}\{\rho(y_{it}, y_{is}, (x_{it} - x_{is})'\beta)\}, \] (5)

with

\[ \rho(y_1, y_2, \delta) = \begin{cases} 
\frac{y_1^2}{2} - \delta y_1 - y_1 y_2, & \text{if } \delta \leq -y_2; \\
(y_1 - y_2 - \delta)^2/2, & \text{if } -y_2 < \delta < y_1; \\
\frac{y_2^2}{2} + \delta y_2 - y_1 y_2, & \text{if } y_1 \leq \delta.
\end{cases} \] (6)

The above objective function satisfies the unconditional moment condition

\[ \frac{\partial A(\beta)}{\partial \beta} = \mathbb{E}\{(e(y_{it} - x_{it}'\beta, x_{is}'\beta) - e(y_{is} - x_{is}'\beta, x_{it}'\beta)) (x_{it} - x_{is})\} = 0, \] (7)
which is a necessary but non-sufficient condition for (4). It gives rise to the nonlinear least squares estimator $\hat{\beta}$ to minimize

$$\sum_{i=1}^{n} \sum_{t<s} \rho(y_{it}, y_{is}, (x_{it} - x_{is})' \beta).$$

(8)

which is the estimator proposed by Honoré (1992) under certain sufficient conditions. Also see Honoré and Kyriazidou (2000), Charlier et al. (2000), Hu (2002), and Honoré and Hu (2004) for related studies.

If $\nu_{it}$ and $\nu_{is}$ are identically distributed conditional on $(x_{it}, x_{is}, \mu_{i})$,

$$E\{\psi(e(y_{it} - x_{it}^t \beta, x_{is}^t \beta)) - \psi(e(y_{is} - x_{is}^t \beta, x_{it}^t \beta))|x_{it}, x_{is}\} = 0$$

(9)

for any function $\psi(\cdot)$. Ai (2005) proposed an estimator that is more efficient than the estimator in Honoré (1992) by utilizing more moment conditions based on the above condition. For integers $k_1$ and $k_2$, let $q(u) = (q_1(u), q_2(u), ..., q_{k_1}(u))'$ denote some known basis functions that approximate any square integrable function of $u$, and let $p(x_{it}, x_{is}) = (p_1(x_{it}, x_{is}), p_2(x_{it}, x_{is}), ..., p_{k_2}(x_{it}, x_{is}))'$ denote some known basis functions that approximate any square integrable function of $(x_{it}, x_{is})$. Equation (9) implies

$$E\{(e(y_{it} - x_{it}^t \beta, x_{is}^t \beta)) - q(e(y_{is} - x_{is}^t \beta, x_{it}^t \beta)) \otimes p(x_{it}, x_{is})\} = 0,$$

(10)

for $t > s$, where $\otimes$ denotes the Kronecker product. Define

$$\phi(y_i, x_i, \beta) = \text{vec}\{(e(y_{it} - x_{it}^t \beta, x_{is}^t \beta)) - q(e(y_{is} - x_{is}^t \beta, x_{it}^t \beta)) \otimes p(x_{it}, x_{is})\}$$

(11)

where $s = 1, 2, ..., T - 1; t = s + 1, s + 2, ..., T$. Then we can estimate the unknown
parameter $\beta$ by minimizing the following objective function in GMM fashion:

$$
\min_{\beta} \left( \sum_{i=1}^{n} \phi(y_i, x_i, \beta) \right)^{'} \hat{\Omega}^{-1} \left( \sum_{i=1}^{n} \phi(y_i, x_i, \beta) \right), \tag{12}
$$

where $\hat{\Omega}^{-1}$ is a positive definite weighting matrix. Ai (2005) showed the asymptotic properties of this GMM estimator by allowing both $k_1$ and $k_2$ to grow with sample size. A disadvantage of the GMM estimator is that the objective function is not globally convex and may have local minimums. One can try many different starting values or simply start with Honoré (2002) estimator.

Along this line of research, Ai and Li (2006) considers the following partially additive semiparametric Tobit model:

$$
y_{it}^{*} = x_{it}^{'} \beta + \sum_{k=1}^{m} h_k(w_{it,k}) + \mu_i + \nu_{it}, \tag{13}
$$

$$
y_{it} = \max\{0, y_{it}^{*}\}, \quad i = 1, 2, \ldots, n; \quad t = 1, 2, \ldots, T, \tag{14}
$$

where $h_k(\cdot)$’s are unknown smooth functions. Using the series method to approximate the unknown functions, Ai and Li (2006) derived the asymptotic properties of the GMM estimator following a procedure similar to (9) to (12). See also Guell and Hu (2006), Chen and Khan (2008), Honoré and Hu (2010) for related studies. More recently, Shiu and Hu (2010) considered nonparametric estimation of nonlinear dynamic panel data models with unobserved covariates. They proposed a sieve maximum likelihood estimator for dynamic discrete choice models and for dynamic censored models and established the asymptotic properties of the estimator.

One difficulty in estimating the Tobit-type panel data models is that the moment conditions or the likelihood functions are highly nonlinear and nontrackable, which often leads to high dimensional integration or other complications. Li and Zheng (2008) considered a
semiparametric Markov Chain Monte Carlo method for dynamic Tobit panel data models. One advantage of this method is that the marginal effects, as well the parameters, can be easily computed.

2.2 Sample selection

The censored models can be viewed as a special sample selection model in which the primary regression and the selection process coincide with each other. Now consider a more general sample selection panel data model:

\[ y_{it} = d_{it} \cdot (x'_{it} \beta + \mu_i + \nu_{it}), \quad (15) \]

\[ d_{it} = 1 \{ z'_{it} \gamma + \alpha_i + \varepsilon_{it} > 0 \}. \quad (16) \]

Conditional on \( d_{it} = 1, d_{is} = 1, x_{it}, z_{it}, x_{is}, z_{is}, \mu_i, \alpha_i \), if \( \nu_{it} \) and \( \nu_{is} \) are identically distributed, we can use the following moment condition to construct a GMM estimator:

\[
E \left\{ \psi(y_{it} - x'_{it} \beta) | d_{it} = 1, d_{is} = 1, x_{it}, z_{it}, x_{is}, z_{is}, \mu_i, \alpha_i \right\} = E \left\{ \psi(y_{is} - x'_{is} \beta) | d_{it} = 1, d_{is} = 1, x_{it}, z_{it}, x_{is}, z_{is}, \mu_i, \alpha_i \right\},
\]

for any function \( \psi(\cdot) \). \( \nu_{it} \) and \( \nu_{is} \) are conditionally identically distributed only if \( z'_{it} \gamma = z'_{is} \gamma \).

Kyazidou (1997) proposed the following kernel estimator

\[
\hat{\beta} = \left[ \sum_i \sum_{s < t} \frac{1}{h} K \left( \frac{(z_{it} - z_{is})' \gamma}{h} \right) d_{it} d_{is} (x_{it} - x_{is}) (x_{it} - x_{is})' \right]^{-1} \left[ \sum_i \sum_{s < t} \frac{1}{h} K \left( \frac{(z_{it} - z_{is})' \gamma}{h} \right) d_{it} d_{is} (x_{it} - x_{is}) (y_{it} - y_{is})' \right], \quad (17)
\]
where $\hat{\gamma}$ is a consistent estimator from (14). Kyiazidou (1997) established the consistency and the asymptotic distribution of the proposed estimator. Honoré and Kyriaizidou (2000b) further extended the model and showed that similar estimators can be constructed for other types of the Tobit models in Amemiya (1985). Again here it is possible to derive more efficient estimators by utilizing more moment conditions similar to that in Ai (2005). Lee and Vella (2006) proposed a semiparametric estimator for censored selection models with endogeneity. Lewbel (2005) proposed a GMM style estimator for panel data models with sample selection. The estimator is based on Honoré and Lewbel (2002), which studies the binary choice panel data model, and Lewbel (2004), which studies selection in cross-sectional models. Semykina and Wooldridge (2010) proposed a parametric correction and a semiparametric correction in panel data models with endogeneity and selection using Mills ratio.

2.3 Attrition and missing Data

The mechanism of attrition and missing data is similar to that of sample selection – some or all variables for some observations have missing values or are not observed. Imbens and Wooldridge (2007) provided a recent survey on the issue. Here we will focus on non/semi-parametric estimations of panel data models with missing data.

Suppose the model is described by the following two equations:

$$y_{it} = d_{it} \cdot (x'_{it} \beta + u_{it}),$$ (18)

$$d_{it} = 1\{z'_{it} \gamma + e_{it} > 0\}.$$ (19)

So when $d_{it} = 0$ the observation on $y_{it}$ and $x'_{it}$ is missing.

Inverse probability weighting (IPW) is often used to estimate models with attrition and missing data. The basic idea of IPW is to weigh the (observed) moment conditions or the loglikelihood function by $\frac{d_{it}}{p_{it}}$, where $p_{it}$ is the propensity of being observed. Robins,
Rotnitzky, and Zhao (1995) proposed a semi-parametric estimator that is based on the IPW principle and a likelihood function that does not have to be fully specified. We maintain the assumption that attrition is an absorbing state – once a person drops out of the sample, she is gone from the sample forever. Suppose the loglikelihood function for the $i$th observation is given by $l_{it}$. The loglikelihood estimation based on the IPW principle can be constructed as

$$
\hat{p}_{it} = \sum_{i=1}^{n} \sum_{t=1}^{T} d_{it} \frac{d_{it}}{\hat{p}_{it}}.
$$

(20)

$p_{it}$ is sequentially estimated as a product of the past probabilities of being in the sample using a binary choice model. Of course such estimators do not have to be based on likelihood functions. Moment conditions can be constructed similarly to equation (20). Rotnitzky, Robins, and Scharfstein (1998) and Vansteelandt, Rotnitzky, and Robins (2007) proposed estimators based on similar principles. Recently Rotnitzky (2008) reviewed the use of IPW in longitudinal data. Moffitt, Fitzgerald, and Gottschalk (1999) was among the first few economists who used the IPW in panel data.

Another approach of dealing with missing data is to impute the missing data. Little and Rubin (2002) provided a comprehensive treatment of missing data, including the imputation methods. However, imputation never seemed to gain much popularity in economics.

Xie, Qian, and Qu (2010) approached the missing data problem from a different angle in cross-sectional models. They proposed a generalized additive model - this allows a wide range of missing data patterns. Their paper essentially proposed the semiparametric version of the index of local sensitivity to nonignorability (Troxel, Ma, and Heitjan, 2004). The proposed method is computationally simple.

Lastly, readers interested in applications of attrition in panel data are referred to a special issue in Journal of Human Resources (Vol. 33.2, 1998) on this topic.
3 Measurement Error

In this section, we review estimation methods for panel data models, which are based on deconvolution and share some similarities with the literature on measurement errors. Measurement errors commonly occur in many economic variables. Carroll et al. (2006) suggest two likely effects of measure error in variables: bias in parameter estimation and loss in power. For recent treatments of measurement errors, see, e.g., Carroll et al (2006) and Chen et al. (forthcoming).

Let $X$ be a covariate of interest. Suppose that $X$ is contaminated by an error $U$ such that the researcher observes instead $W = X + U$. In this case, a naive regression of $Y$ on the observed $W$, as a proxy for $X$, usually gives biased estimate of $E[Y|X]$. For instance, consider the simple linear regression $Y = \beta_0 + \beta X + \nu$, where $E[\nu|X] = 0$. Under the assumption of classical measurement error that $U$ is independent of $X$ and $\nu$, and $E[U] = 0$, the OLS coefficient of $Y$ on $W$ converges to $\lambda \beta$, where $\lambda = \frac{\sigma^2}{\sigma^2_u + \sigma^2_\nu}$ and $\sigma^2_\nu$ and $\sigma^2_u$ are the variance of $\nu$ and $u$ respectively. This is the well known attenuation bias of classical measurement errors in linear regressions.

The bias due to measurement errors can be more complicated than the simple attenuation bias, depending on the nature of the measurement error and the functional form of the model in question. There is a large body of literature on the treatment of measurement errors, with a focus on removing or mitigating the bias. Usually validation data are required to tackle measure error issues. Depending on the extent to which parametric assumptions are utilized, various methods can be grouped into three broad categories: parametric, semi-parametric and nonparametric methods.

In this subsection, we focus on the nonparametric methods for measurement errors. A common theme in nonparametric treatments of measurement error is the method of deconvolution. For the example given above, $W$ is the result of convolution between two random
variables $X$ and $U$. Consequently, the estimation of the distribution of $X$ given that of $U$ and data on $W$ is called deconvolution.

We note that a panel data model can be treated in the framework of measure error problem in the sense that the individual effect $\mu_i$ can be viewed as a measurement error to the (unobserved) conditional mean $E[y_{it}|x_{it}]$. As a matter of fact, this is the approach employed by Horowitz and Makartou (1996) and Evdokimov (2010), among others.

Below we first review the deconvolution of densities, followed by a detailed discussion of a deconvolution-based estimator of panel data model by Horowitz and Markatou (1996) and further developments offered in Evdokimov (2010).

### 3.1 Deconvolution of densities

The building stocks of nonparametric deconvolution are the characteristic functions. See, e.g., Stuart and Ord (1987) for detailed treatments. The characteristic function of a random variable $X$ with density $f_X$ is given by $\psi_x(t) = \int \exp(itx)f_X(x)dx$, where $i = \sqrt{-1}$. Assume that the measurement error $U$ is independent of $X$ and $\psi_u(t) \neq 0$ for all real, finite $t$, then $\psi_W(t) = \psi_X(t)\psi_U(t)$. It follows that one can obtain the density of $X$ through the inverse Fourier transformation:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} \psi_W(t)}{\psi_U(t)} dt. \tag{21}$$

Suppose the distribution of $U$ is known, deconvolution proceeds by first estimating $\hat{\psi}_W$ and then calculating the density of $X$ via the inverse Fourier transformation. Like the empirical CDF function, the characteristic function of $W$ can be estimated as the following:

$$\hat{\psi}_W(t) = \frac{1}{n} \sum_{i=1}^{n} \exp(itW_i).$$

---

1 We use $i$ rather than the usual $i$ for the imaginary unit to distinguish it from the typical subscript $i$ used in panel data models.
One can then replace $\psi_W$ in the deconvolution formula by $\hat{\psi}_W$ to estimate $f_X$. There is, however, one numerical difficulty associated with this approach. The inverse Fourier transformation (21) assumes that the distribution of $W$ has a density such that $\psi_W(t)/\psi_U(t) \to 0$ as $t \to \infty$. But the empirical distribution of $W$ is discrete and has no density. Thus $\hat{\psi}_W(t)/\psi_U(t)$ does not necessarily converge to 0 as $t \to \infty$. A method to solve this problem is to convolute the empirical distribution of $W$ with the distribution of a continuously distributed random variable that becomes degenerate as $n \to \infty$.

Smoothing the distribution of $W$ amounts to multiplying the empirical characteristic function $\hat{\psi}_W$ by the characteristic function of a smooth distribution. In particular, let $\zeta$ be a random variable with a bounded characteristic function $\psi_\zeta$ whose support is $[-1,1]$. Further let $\{v_n\}$ be a sequence of positive constants that converges to 0 as $n \to \infty$. The smoothing is carried out by replacing $W$ with $W + v_n\zeta$ in the estimation, yielding the following smoothed/regularized deconvolution estimator of $f_X$:

$$
\hat{f}_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\hat{\psi}_W(t)\psi_\zeta(v_n\zeta)}{\psi_U(t)} dt. \tag{22}
$$

The integral (22) exists because the integral is 0 if $|t| > 1/v_n$. If $v_n$ does not converge to 0 too rapidly as $n \to \infty$, then $\hat{f}_X(x)$ is consistent for $f_X(x)$.

If the distribution and subsequently characteristic function of $U$ is unknown, one can in principle replace it by its sample counterpart. It follows that the density of $X$ is estimated by

$$
\hat{f}_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \frac{\hat{\psi}_W(t)\psi_\zeta(v_n\zeta)}{\psi_U(t)} dt. \tag{23}
$$

Note that only one regularization term is used when either or both characteristic functions of $W$ and $U$ are replaced by their sample counterparts. Lastly the CDF of $X$ can be estimated
by integrating $\hat{f}_X$:

$$\hat{F}_X(x) = \int_{-M_n}^{x} \hat{f}_X(z) dz,$$

where $M_n \to \infty$ as $n \to \infty$.

The convergence rate of deconvolution is often very slow. Carroll and Hall (1988), Fan (1991), and Stefanski and Carroll (1990, 1991) investigate the asymptotic properties of $\hat{f}_X$. Under some regularity conditions, $\hat{f}_X$ converges to $f_X$ at the rate in power of $\ln n$, which is slower than the root-$n$ rate of parametric estimators and the usual $n^{-2/5}$ rate for nonparametric density estimators. It transpires that the rate of convergence is largely controlled by the thickness of the tails of the characteristic functions of $U$, which in turn is determined by the smoothness of the distribution of $U$. For example, when the tails of characteristic functions of $U$ decrease exponentially fast, the fastest possible rate of convergence is logarithmic of $n$. This family includes the normal, Cauchy, and Type I extreme value distributions. Faster convergence rates are possible when the tails of the characteristic functions of $U$ decrease only geometrically fast. The Laplace and symmetrical gamma distribution satisfy this condition. For a detailed discussion on the large sample properties of deconvolution estimation, see Horowitz (2009).

### 3.2 Linear panel models

The method of deconvolution can be used in panel data model in which a composite error can be viewed as the convolution between the individual effect and an idiosyncratic error term. This approach is taken by Horowitz and Markatou (1996), which is reviewed below.

Consider again the following typical linear panel model:

$$y_{it} = x_{it}'\beta + \mu_i + \nu_{it}, i = 1, \ldots, n.$$
For simplicity, let $t = 1, 2$. Let $\hat{\beta}$ be an $n^{-1/2}$-consistent estimator of $\beta$. Define, for $i = 1, \ldots, n$,

$$W_{it} = Y_{it} - X'_{it}\hat{\beta}, \ t = 1, 2,$$

$$\eta_i = W_{i1} - W_{i2}.$$ 

It follows that as $n \to \infty$, $W_{it}$ converges in distribution to $W = \mu + \nu$ and $\eta_i$ converges in distribution to the random variable $\eta$ that is distributed as the difference between two independent realizations of $\nu$. It is seen that if the distribution of $\nu$ is known, one can use the deconvolution estimator discussed above to estimate the density of $\mu$. The problem here, however, is that the distribution of $\nu$ is not known; furthermore, it cannot be sampled directly since we do not observe random realizations of $\nu$, which are convoluted with the individual effect in panel models.

Although $\psi_\nu$ cannot be estimated directly, $\psi_\eta$ can be estimated from the first differenced errors. Note that $\psi_\eta(t) = |\psi_\nu(t)|^2$, where $| \cdot |$ denotes the modulus of the complex variable between the bars. Horowitz and Markatou (1996) proceed by assuming that the distribution of $\nu$ is symmetric about zero and $\psi_\nu(t) \neq 0$ for all finite $t$. It follows that $\psi_\nu(t)$ is real and strictly positive for all finite $t$, and

$$\psi_\nu(t) = \psi_\eta(t)^{1/2}.$$ 

We can then estimate the density of $\nu$ and $\mu$ as follows:

$$f_\nu(\nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\nu} \psi_\eta(t)^{1/2} dt,$$

$$f_\mu(\mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\mu} \frac{\psi_W(t)}{\psi_\eta(t)^{1/2}} dt.$$
Since $\psi_W$ and $\psi_\eta$ are unknown, they are estimated by their sample analogs:

\[
\hat{\psi}_W(t) = \frac{1}{2n} \sum_{i=1}^{n} \{\exp(itW_{1i}) + \exp(itW_{2i})\}
\]

\[
\hat{\psi}_\eta(t) = \frac{1}{n} \sum_{i=1}^{n} \exp(it\eta_i).
\]

As discussed above, $f_\nu$ and $f_\mu$ may not be estimated if we simply plug the empirical characteristic functions $\hat{\psi}_W$ and $\hat{\psi}_\eta$ into the inverse Fourier transformation formula because the resulting integrals do not exist in general. Again regularization is called for to smooth the empirical distributions of $W$ and $\eta$. Let $\zeta$ be a random variable whose characteristic function $\psi_\zeta$ has support $[-1,1]$ and $\{v_{n\nu}\}$ and $\{v_{n\mu}\}$ be sequences of positive bandwidths that converge to 0 as $n \to \infty$. The smoothed estimator of $f_W$ and $f_\nu$ are given by

\[
\hat{f}_\nu(\nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\nu} |\hat{\psi}_\eta(t)|^{1/2} \psi_\zeta(v_{n\nu}t) dt
\]

\[
\hat{f}_\mu(\mu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\mu} \frac{\hat{\psi}_W(t) \psi_\zeta(v_{n\mu}t)}{|\hat{\psi}_\eta(t)|^{1/2}} dt.
\]

Horowitz and Markatou (1996) establish the consistency, convergence rates and asymptotic normality of $\hat{f}_\nu$ and $\hat{f}_\mu$ under some regularity conditions. They also show that the assumption that $\nu$ is symmetrically distributed can be relaxed and propose a method to estimate the distribution of $\nu$ without the symmetry assumption. In addition, they present a method to deal with serially correlated idiosyncratic errors.

Although the distribution of $\nu$ and $\mu$ are often of little direct interest, knowledge of them are useful in many economic applications. Horowitz and Markatou (1996) use their proposed method to estimate an earning model using data from the Current Population Survey. The model’s transitory error component is shown to be not normally distributed. Use of the deconvolution based nonparametric estimators yields estimates of the probability
that individuals with low earnings will become high earners in the future that are much lower than the estimates obtained under the assumption of normally distributed error components.

Lastly, Horowitz and Markatou (1996) show in their Monte Carlo simulations that the deconvolution estimators can have large biases, which affect the final estimates considerably. They propose a method of bias reduction to remove biases caused by smoothing of the empirical distribution of $\eta$. Substantial improvements in the estimated transition probabilities are reported after this bias reduction.

### 3.3 Nonparametric models for panel data

Evdokimov (2010) generalizes the method proposed by Horowtiz and Markatou (1996) in two directions. He considers a nonparametric panel data model with nonadditive unobserved heterogeneity. The individual specific effects enter the structural function nonseparably and are allowed to be correlated with the covariates in an arbitrary manner. The idiosyncratic disturbance term is additively separable from the structural function. Nonparametric identification of all the structural elements of the model is established. No parametric distributional or functional form assumptions are needed for identification. Thus, the model permits nonparametric distributional and counterfactual analysis of heterogeneous marginal effects using short panels.

In particular, Evdokimov (2010) considers the following model

$$ y_{it} = m(x_{it}, \mu_i) + \nu_{it}, i = 1, \ldots, n, t = 1, \ldots, T, $$

where $m(x_{it}, \mu_i)$ is the conditional mean of $y_{it}$ given observed covariates $x_{it}$ and unobserved individual effect $\mu_i$. No functional assumptions are imposed except that $m$ is increasing in $\mu_i$. The identification of the structural function $m(x_{it}, \mu_i)$ in Evdokimov (2010) is based on the following Lemma, which is an extension of Kotlarski (1967).
Lemma 1 (Evdokimov, 2010) Suppose \((Y_1, Y_2) = (A + U_1, A + U_2)\), where the scalar random variables \(A, U_1\) and \(U_2\), (i) are mutually independent, (ii) have at least one absolute moment, (iii) \(E[U_1] = 0\), (iv) \(\psi_{U_1}(s) \neq 0\) for all \(s\) and \(t \in \{1, 2\}\). Then, the distribution of \(A, U_1\) and \(U_2\) are identified from the joint distribution of \((Y_1, Y_2)\).

Evdokimov (2010) shows that under some regularity conditions, the panel model (24) is identified for a panel with at least two periods. The estimator is constructed in three steps. First, the conditional (on covariates \(X_{it}\)) distribution of the idiosyncratic disturbances is identified using the information on the subset of individuals whose covariates do not change across time periods. Next, conditional on covariates, one can deconvolve \(\nu_{it}\) from \(Y_{it}\) to obtain the conditional distribution of \(m(X_{it}, \mu_i)\) that is the key to identifying the structural function. The third step identifies the structural function under two different scenarios corresponding to the random effects or the fixed effects in the usual panel data model estimations. To save space, below we focus our discussion on the random effects model, which assumes that the individual effects and the observed covaraites are independent. It is also assumed that the individual effect \(\mu_i\) has a standard uniform distribution — a normalization that is necessary since the function \(m(x, \mu)\) is modeled nonparametrically.

The key component of Evdokimov (2010)’s estimators is the conditional deconvolution. When the covariates are discrete, estimation can be performed using the existing deconvolution techniques. The sample us split into subgroups according to the values of the covariates and a deconvolution is be used to obtain the estimates of necessary cumulative distribution functions. In contrast, when the covariates are continuously distributed one needs to use a conditional deconvolution procedure. Evdokimov (2010) investigates the estimation and large sample properties of conditional characteristic function estimations, which not only plays a key role in his nonparametric estimators of panel data models, but also is of independent interest in itself.

Noting that the conditional characteristic function \(\psi_{Y_{it}}(s|X_{it}) = E[\exp(\iota s Y_{it})|X_{it} = x_{it}]\),
Evdokimov (2010) proposes a kernel estimator

\[ \hat{\psi}_{Y_{it}}(s|x_t) = \frac{\sum_{i=1}^{n} \exp(\imath s Y_{it}) K_{h_Y}(X_{it} - x_t)}{\sum_{i=1}^{n} K_{h_Y}(X_{it} - x_t)}, \]

where \( h_Y \to 0 \) is a bandwidth parameter, \( K_h(\cdot) = K(\cdot/h)/h \) and \( K(\cdot) \) is a standard kernel function. Under the random effect model assumption that \( \mu_i \) and \( X_{it} \) are independent, the distributions of the idiosyncratic errors are identified based on Lemma 1. Their corresponding conditional distributions can be obtained through conditional deconvolutions in a similar manner. Consider first the simplest case where the conditional distribution of \( \nu_{it} \) is symmetric and the same across \( t = 1, 2 \). Conditional on \( X_{i1} = X_{i2} = x \) for all \( x \in \mathcal{X} \), then for any \( x \in \mathcal{X} \), the conditional characteristic function of \( Y_{i2} - Y_{i1} \) equals

\[ \psi_{Y_{i2} - Y_{i1}}(s|X_{i2} = X_{i1} = x) = \psi_{\nu}(s|x)\psi_{\nu}(-s|x) = |\psi_{\nu}(s|x)|^2, \]

where \( \psi_{\nu}(s|x) \) is the conditional characteristic function of \( \nu_{it} \), conditional on \( X_{it} = x \). It follows that this conditional characteristic function can be estimated by

\[ \hat{\psi}_{\nu}(s|x) = \left[ \frac{\sum_{t=1}^{T-1} \sum_{\tau=t+1}^{T} \sum_{i=1}^{n} \exp\{\imath s(Y_{it} - Y_{i\tau})\} K_{h_{\nu}}(X_{it} - x)K_{h_{\nu}}(X_{i\tau} - x)}{\sum_{t=1}^{T-1} \sum_{\tau=t+1}^{T} \sum_{i=1}^{n} K_{h_{\nu}}(X_{it} - x)K_{h_{\nu}}(X_{i\tau} - x)} \right]^{1/2}, \]

where the bandwidth \( h_{\nu} \to 0 \). In the absence of the assumption that the distribution of \( \nu_{it} \) is symmetric and remains the same across the time periods, one can estimate the conditional
characteristic function of \( \nu_{it} \) using the following general formula\(^2\)

\[
\hat{\psi}_{\nu_{it}}(s|x_t) = \frac{1}{T-1} \sum_{\tau=1}^{T} \exp \left\{ \tau \int_{0}^{s} \sum_{i=1}^{n} \frac{Y_{it} e^{\tau(Y_{it} - Y_{i\tau})} K_{h_{\nu}}(X_{it} - x) K_{h_{\nu}}(X_{i\tau} - x)}{\sum_{i=1}^{n} e^{\tau(Y_{it} - Y_{i\tau})} K_{h_{\nu}}(X_{it} - x) K_{h_{\nu}}(X_{i\tau} - x)} \, d\xi \right. \\
- \left. \tau s \sum_{i=1}^{n} Y_{it} K_{h_{\nu}}(X_{it} - x_t) \right\}.
\]

Next using the fact that

\[
\psi_{m(X_{it}, \mu_{i})}(s|X_{it} = x_t) = \psi_{Y_{it}}(s|X_{it} = x_t)/\psi_{\nu_{it}}(s|X_{it} = x_t),
\]

we can estimate the conditional characteristic function using its empirical analog \( \hat{\psi}_{m(x_t, \mu_i)} = \hat{\psi}_{Y_{it}}(s|x_t)/\hat{\psi}_{\nu_{it}}(s|x_t) \). One can then calculate a smoothed conditional CDF of \( m(x_t, \mu_i) \) using the following formula:

\[
\hat{F}_{m(x_t, \mu_i)}(w|x_t) = \frac{1}{2} - \int_{-\infty}^{w} \frac{e^{-i\xi w}}{2\pi i \xi} \psi_\zeta(h_\zeta \xi) \hat{\psi}_{m(x_t, \mu_i)}(s|x_t) \, ds,
\]

where \( \psi_\zeta(\cdot) \) is the Fourier transform of a kernel function \( \zeta(\cdot) \) and \( h_\zeta \rightarrow 0 \) is a bandwidth. Again, this extra convolution is needed to ensure integrability of the empirical conditional characteristic function. Since the estimated CDF obtained by inverting an empirical characteristic function is not necessarily monotonic, the rearrangement technique in Chernozhukov et al. (2007) is employed to render the estimated CDFs monotonic and the estimated quantile functions well behaved. It is noted that the random effect model estimated in this manner can be interpreted as quantile regression with measurement error \( \nu_{it} \).

Denote the rearranged version of the estimated conditional CDF by \( \tilde{F}_{m(x_t, \mu_i)}(w|x_t) \). The

\(^2\)See the proof of Lemma 1 in Evdokimov (2010) for the derivation of this general result.
corresponding conditional quantile function is then given by

$$\hat{Q}_{m(x_t, \mu_i)}(q| x_t) = \min_w \left\{ \hat{F}_{m(x_t, \mu_i)}(w|x_t) \geq q \right\}. $$

Finally, the structural function of interest in the random effect model is given by

$$\hat{m}(x, \mu) = \frac{1}{T} \sum_{t=1}^{T} \hat{Q}_{m(x, \mu)}(\mu| X_{it} = x), \mu \in (0, 1).$$

Evdokimov (2010) then proceeds to establish the large sample properties of the proposed estimator. An alternative fixed effect estimator is also presented. Interested readers are referred to his paper for details. A small scale Monte Carlo simulations illustrates the finite sample performance of the the proposed estimators.

The deconvolution based method for panel data is also employed by Bonhomme and Robin (2008) in a multi-factor loading model. They construct a non-parametric estimator of the distributions of latent factors in linear independent multi-factor models under the assumption that factor loadings are known. Their approach allows estimation of the distributions of up to \(L(L+1)/2\) factors given \(L\) measurements. Their estimator uses empirical characteristic functions, like many available deconvolution estimators. They apply the generalized deconvolution procedure to decompose individual log earnings from the panel study of income dynamics (PSID) into permanent and transitory components.

We conclude this section by referring interested readers to the recent developments in the literature on measurement errors that use deconvolution methods. Many of these estimators can be applied to panel data model whose error components consist of an individual specific effect and an idiosyncratic error. The list includes, among others, Li (2002), Hong and Tamer (2003), Schennach (2004a, 2004b, 2007), Hu and Schennach (2008) and Chen, Hu, and Lewbel (2008a, 2008b, 2009). In addition, Zhang (2002) and Kim (2007) suggest some
semiparametric estimation strategies for panel count data with measurement errors. Bjørn and Krishnakumar (2008) present a survey on recent measurement error in panel data.

4 Concluding remarks

In this article we review some recent developments in the semi- or non-parametric estimations of panel data models, for which some variables are only partially observed. We focus on two areas of interest: (i) panel models with discrete/limited dependent variables; (ii) estimations of panel models based on nonparametric deconvolution methods. Due to the limitation of time and space, a comprehensive coverage of the rapidly growth literature on the semi-parametric and non-parametric estimations of panel data models is beyong the scope of this review. Interested readers are referred to Diggle et al. (2002), Arellano (2003), Arellano and Hahn (2007), and Li and Racine (2007) for general treatments on the recent developments of panel data models.

References


