

Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity*

Kirill Evdokimov[†]

Department of Economics
Princeton University

August, 2010

Abstract

This paper considers a nonparametric panel data model with nonadditive unobserved heterogeneity. As in the standard linear panel data model, two types of unobservables are present in the model: individual-specific effects and idiosyncratic disturbances. 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. The identification result is constructive and only requires panel data with two time periods. Thus, the model permits nonparametric distributional and counterfactual analysis of heterogeneous marginal effects using short panels. The paper also develops a nonparametric estimation procedure and derives its rate of convergence. As a by-product the rates of convergence for the problem of conditional deconvolution are obtained. The proposed estimator is easy to compute and does not require numeric optimization. A Monte-Carlo study indicates that the estimator performs very well in finite samples.

*This paper is a revision of the first chapter of my thesis. I am very grateful to Donald Andrews, Xiaohong Chen, Yuichi Kitamura, Peter Phillips, and Edward Vytlačil for their advice, support, and enthusiasm. I am especially thankful to my advisor Yuichi Kitamura for all the effort and time he spent nurturing me intellectually. I have also benefited from discussions with Joseph Altonji, Stephane Bonhomme, Martin Browning, Victor Chernozhukov, Flavio Cunha, Bryan Graham, Jerry Hausman, James Heckman, Stefan Hoderlein, Joel Horowitz, Ilze Kalnina, Lung-Fei Lee, Simon Lee, Yoonseok Lee, Taisuke Otsu, James Powell, Pavel Stetsenko, and Quang Vuong. I also thank seminar participants at Yale, Brown, Chicago Booth, Chicago Economics, Duke, EUI, MIT/Harvard, Northwestern, Princeton, Toulouse, UC Berkeley, UCLA, Warwick, North American Summer Meeting of the Econometric Society 2009, SITE 2009 Summer Workshop on Advances in Nonparametric Econometrics, and Stats in the Chateau Summer School for very helpful comments. All errors are mine. Financial support of Cowles Foundation via Carl Arvid Anderson Prize is gratefully acknowledged.

[†]E-mail: kevdokim@princeton.edu.

1 Introduction

The importance of unobserved heterogeneity in modeling economic behavior is widely recognized. Panel data offer useful opportunities for taking latent characteristics of individuals into account. This paper considers a nonparametric panel data framework that allows for heterogeneous marginal effects. As will be demonstrated, the model permits nonparametric identification and estimation of all the structural elements using short panels.

Consider the following panel data model:

$$Y_{it} = m(X_{it}, \alpha_i) + U_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T; \quad (1)$$

where X_{it} is a vector of explanatory variables, Y_{it} is a scalar outcome variable, scalar α_i represents persistent heterogeneity (possibly correlated with X_{it}), and U_{it} is a scalar idiosyncratic disturbance term.^{1,2} This paper assumes that the number of time periods T is *small* ($T = 2$ is sufficient for identification), while the number of cross-section units n is large, which is typical for microeconomic data.

This paper explains how to nonparametrically identify and estimate the *structural function* $m(x, \alpha)$ unknown to the econometrician. The paper also identifies and estimates the conditional distribution of α_i , given X_{it} , which is needed for policy analysis. The analysis does not impose any parametric assumptions on the function $m(x, \alpha)$ or on the distributions of α_i and U_{it} .

The structural function depends nonlinearly on α , which allows the derivative $\partial m(x, \alpha) / \partial x$ to vary across units with the same observed x . That is, observationally identical individuals can have different responses to changes in x . This is known to be an important feature of microeconomic data.³ This paper shows, among other things, how to estimate the distribution of heterogeneous marginal effects fully nonparametrically.

As an example of application of the above model, consider the effect of union membership on wages. Let Y_{it} denote the (logarithm of) individual's wage and X_{it} be a dummy coding whether the i -th individual wage in t -th period was negotiated as a part of union agreement. Heterogeneity α_i is the unobserved individual skill level, while the idiosyncratic disturbances U_{it} denote luck and measurement error. The effect of union membership for an individual with

¹As usual, capital letters denote random variables, while lower case letters refer to the values of random variables. The only exception from this rule is α_i , which is a random variable, while α stands for its value. This notation should not cause any confusion because α_i is an unobservable.

²It is assumed that U_{it} is independent of α_i , conditional on X_{it} . The disturbance U_{it} does not have to be independent of X_{it} ; only the standard orthogonality restriction $E[U_{it}|X_{it}] = 0$ is imposed. In addition, the distribution of U_{it} does not have to be the same across time periods.

³A partial list of related studies includes Heckman, Smith, and Clements (1997), Heckman and Vytlačil (1998), Abadie, Angrist, and Imbens (2002), Matzkin (2003), Chesher (2003), Chernozhukov and Hansen (2005), Imbens and Newey (In press), Bitler, Gelbach, and Hoynes (2006), Djebbari and Smith (2008), and the references therein.

skill α_i is then $m(1, \alpha_i) - m(0, \alpha_i)$. Nonseparability of the structural function in α permits the union membership effect to be different across individuals with different unobserved skill levels. This is exactly what the empirical literature suggests.⁴ On the contrary, a model with additively separable α_i (*e.g.*, linear model $m(X_{it}, \alpha_i) = X'_{it}\beta + \alpha_i$) fails to capture the heterogeneity of the union effect because α_i cancels out.

Similar to the above example, the model can be applied to the estimation of treatment effects when panel data are available. Assumptions imposed in this model are strong, but allow nonparametric identification and estimation of heterogeneous treatment effects for both treated and untreated individuals. Thus, the model can be used to study the effect of a policy on the whole population. For instance, one can identify and estimate the percentage of individuals who are better/worse off as a result of the program.

Another example is the life-cycle model of consumption and labor supply of Heckman and MaCurdy (1980) and MaCurdy (1981).⁵ These papers obtain individual consumption behavior of the form⁶

$$C_{it} = \underbrace{C(W_{it}, \lambda_i)}_{\text{unknown function}} + \underbrace{U_{it}}_{\text{measurement error}} \quad (2)$$

where C_{it} is the (logarithm of) consumption, W_{it} is the hourly wage, U_{it} is the measurement error, and λ_i is the scalar unobserved heterogeneity that summarizes all the information about individual's initial wealth, expected future earnings, and the form of utility function. The consumption function $C(w, \lambda)$ can be shown to be increasing in both arguments, but is unknown to the researcher since it depends on the utility function of the individual. In the existing literature, it is common to assume very specific parametric forms of utility functions for estimation. The goal of these parametric assumptions is to make the (logarithm of) consumption function additively separable in the unobserved λ_i . Needless to say, this may lead to model misspecification. In contrast, the method of this paper can be used to estimate model (2) without imposing any parametric assumptions.⁷

The nonparametric identification result of this paper is constructive and only requires two periods of data ($T = 2$). The identification strategy consists of the following three steps. First, the conditional (on covariates X_{it}) distribution of the idiosyncratic disturbances U_{it} 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 U_{it} from Y_{it} to obtain the conditional distribution of $m(X_{it}, \alpha_i)$ that is the key to identifying the structural function. The third step identifies the structural function under two different scenarios. One scenario assumes random effects, that is, the unobserved heterogeneity is independent of the

⁴See, for example, Card (1996), Lemieux (1998), and Card, Lemieux, and Riddell (2004).

⁵I thank James Heckman for suggesting the example.

⁶I have imposed an extra assumption that the rate of intertemporal substitution equals the interest rate.

⁷Similarly, restrictive parametric forms of utility functions are imposed in studies of family risk-sharing and altruism, see for example Hayashi, Altonji, and Kotlikoff (1996) and the references therein.

covariates. The other scenario considers fixed effects, so that the unobserved heterogeneity and the covariates could be dependent.⁸ The latter case is handled without imposing any parametric assumptions on the form of the dependence. Similar to linear panel data models, in the random effects case between-variation identifies the unknown structural function. In the fixed effects case one has to rely on within-variation for identification of the model. The details of the identification strategy are given in Section 2.

The paper proposes an estimation procedure that is easy to implement. No numerical optimization is necessary. Estimation of the model boils down to estimation of conditional cumulative distribution and quantile functions. Conditional cumulative distribution functions (conditional CDFs) are estimated by conditional deconvolution and quantile functions are estimated by inverting the corresponding conditional CDFs. Although deconvolution has been widely studied in the statistical and econometric literature, this paper appears to be the first to consider conditional deconvolution. The paper provides the necessary estimators of conditional CDFs and derives their rates of convergence. The estimators, assumptions, and theoretical results are presented in Section 3. The rates of convergence of the conditional deconvolution estimators of conditional CDFs are shown to be natural combinations of the rates of convergence of the unconditional deconvolution estimator (Fan, 1991) and the conditional density estimator (Stone, 1982).⁹ Finite sample properties of the estimators are investigated by a Monte-Carlo study in Section 4. The estimators appear to perform very well in practice.

The literature on parametric and semiparametric panel data modelling is vast. Traditional linear panel models with heterogeneous intercepts are reviewed, for example, in Hsiao (2003) and Wooldridge (2002). Hsiao (2003) and Hsiao and Pesaran (2004) review linear panel models with random individual slope coefficients. Several recent papers consider fixed effect estimation of nonlinear (semi-)parametrically specified panel models. The estimators for the parameters are biased, but large T asymptotic approximations are used to reduce the order of bias; see for example Arellano and Hahn (2006) or Hahn and Newey (2004). As an alternative, Honoré and Tamer (2006) and Chernozhukov, Fernandez-Val, Hahn, and Newey (2008) consider set identification of the parameters and marginal effects.

A line of literature initiated by Porter (1996) studies panel models, where the effect of covariates is not restricted by a parametric model. However, heterogeneity is still modeled as an additively separable intercept, i.e. $Y_{it} = g(X_{it}) + \alpha_i + U_{it}$. See, for example, Henderson, Carroll, and Li (2008) for a list of references and discussion. Since $g(\cdot)$ is estimated nonparametrically, these models are sometimes called nonparametric panel data models. The analysis of this paper is "more" nonparametric since it models the effect of heterogeneity α_i fully nonparametrically.

⁸See Graham and Powell (2008) for the discussion of the "fixed effects" terminology.

⁹The conditional deconvolution estimators and the statistical results of Section 3 may also be used in other applications such as measurement error or auction models.

Conceptually, this paper is related to the work of Kitamura (2004), who considers non-parametric identification and estimation of a finite mixture of regression functions using cross-section data. Model (1) can be seen as an infinite mixture model. The assumption of finiteness of the number of mixture components is crucial for the analysis of Kitamura (2004), and his assumptions, as well as his identification and estimation strategies are different from those used here. The method of this paper is also related to the work of Horowitz and Markatou (1996). They study the standard linear random effect panel, but do not impose parametric assumptions on the distribution of either the heterogeneous intercept or idiosyncratic errors.

For fully nonseparable panel models, Altonji and Matzkin (2005) and Bester and Hansen (2007) present conditions for identification and estimation of the local average derivative. Their results do not identify the structural function, policy effects or weighted average derivative in model (1). Altonji and Matzkin (2005) and Athey and Imbens (2006) nonparametrically identify distributional effects in panel data and repeated cross-section models with scalar endogenous unobservables. In contrast, model (1) clearly separates the role of unobserved heterogeneity and idiosyncratic disturbances.

Chamberlain (1992) considers a linear panel data model with random coefficients, independent of the covariates. Lemieux (1998) considers the estimation of a linear panel model where fixed effects can be interacted with a binary covariate (union membership). His assumption and model permit variance decompositions that are used to study the effect of union membership on wage inequality. Note that model of this paper does not impose linearity and allows for complete distributional analysis. In particular, the model can be applied to study how the individual's unobserved characteristics affect his or her likelihood of becoming a union member; a result that cannot be obtained from a variance decomposition.

Recently, Arellano and Bonhomme (2008) and Graham and Powell (2008) consider the analysis of linear panel models where the coefficients are random and can be arbitrarily correlated with the covariates. The first paper identifies the joint distribution of the coefficients, while the second paper obtains average partial effect under weaker assumptions. While developed independently of the current paper, Arellano and Bonhomme (2008) also use deconvolution arguments for identification. At the same time, there are at least two important differences between the approach of the present paper and the models of Arellano and Bonhomme (2008) and Graham and Powell (2008). The linearity assumption is vital for the analysis of both papers. In addition, the identifying rank restriction imposed in these papers does not allow identification of some important counterfactuals such as the treatment effect for the untreated or for the whole population. The present paper does not rely on linearity for identification and does provide identification results for the above-mentioned counterfactuals. Finally, the very recent papers by Chernozhukov, Fernandez-Val, and Newey (2009), Graham, Hahn, and Powell (2009), and Hoderlein and White (2009) are also related to the present paper, but have a different focus.

2 Identification

This section presents the identification results for model (1). The identification results are presented for the case $T = 2$; generalization for the case $T > 2$ is immediate. Consider the following assumption:

Assumption ID. *Suppose that:*

- (i) $T = 2$ and $\{X_i, U_i, \alpha_i\}_{i=1}^n$ is a random sample, where $X_i \equiv (X_{i1}, X_{i2})$ and $U_i \equiv (U_{i1}, U_{i2})$;
- (ii) $f_{U_{it}|X_{it}, \alpha_i, X_{i(-t)}, U_{i(-t)}}(u_t|x, \alpha, x_{(-t)}, u_{(-t)}) = f_{U_{it}|X_{it}}(u_t|x)$ for all $(u_t, x, \alpha, x_{(-t)}, u_{(-t)}) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X} \times \mathbb{R}$ and $t \in \{1, 2\}$;¹⁰
- (iii) $E[U_{it}|X_{it} = x] = 0$, for all $x \in \mathcal{X}$ and $t \in \{1, 2\}$;¹¹
- (iv) the (conditional) characteristic function $\phi_{U_{it}}(s|X_{it} = x)$ of U_{it} , given $X_{it} = x$, does not vanish for all $s \in \mathbb{R}$, $x \in \mathcal{X}$, and $t \in \{1, 2\}$;¹²
- (v) $E[|m(x_t, \alpha_i)| | X_i = (x_1, x_2)]$ and $E[|U_{it}| | X_{it} = x_t]$ are uniformly bounded for all t and $(x_1, x_2) \in \mathcal{X} \times \mathcal{X}$;
- (vi) the joint density of (X_{i1}, X_{i2}) satisfies $f_{X_{i1}, X_{i2}}(x, x) > 0$ for all $x \in \mathcal{X}$, where for discrete components of X_{it} the density is taken with respect to the counting measure;
- (vii) $m(x, \alpha)$ is weakly increasing in α for all $x \in \mathcal{X}$;
- (viii) α_i is continuously distributed, conditional on $X_i = (x_1, x_2)$, for all $(x_1, x_2) \in \mathcal{X} \times \mathcal{X}$
- (ix) functions $m(x, \alpha)$, $f_{U_{it}|X_{it}}(u|x)$, $f_{\alpha_i|X_{it}}(\alpha|x)$, $f_{\alpha_i|X_{i1}, X_{i2}}(\alpha|x_1, x_2)$ are everywhere continuous in the continuously distributed components of x , x_1 , x_2 for all $\alpha \in \mathbb{R}$ and $u \in \mathbb{R}$.¹³

Conditional independence Assumption ID(ii) is strong; however, independence assumptions are usually necessary to identify nonlinear nonparametric models. For example, this assumption is satisfied if $U_{it} = \sigma_t(X_{it})\xi_{it}$, where $\sigma_t(x)$ are positive bounded functions and ξ_{it} are *i.i.d.* $(0, 1)$ and are independent of (α_i, X_i) . Assumption ID(ii) rules out lagged dependent variables as explanatory variables, as well as serially correlated disturbances. On the other hand, this assumption permits conditional heteroskedasticity, since U_{it} does not need

¹⁰Index $(-t)$ stands for "other than t " time periods, which in the case $T = 2$ is the $(3 - t)$ -th period.

¹¹It is possible to replace Assumption ID(iii) with the conditional quantile restriction $Q_{U_{it}|X_{it}}(q|x) = 0$ for some $q \in (0, 1)$ for all $x \in \mathcal{X}$.

¹²The conditional characteristic function $\phi_A(s|X = x)$ of A , given $X = x$, is defined as $\phi_A(s|X = x) = E[\exp(isA) | X = x]$, where $i = \sqrt{-1}$.

¹³It is straightforward to allow the functions to be *almost* everywhere continuous. In this case the function $m(x, \alpha)$ is identified at all points of continuity.

to be independent of X_{it} . Moreover, the conditional and unconditional distributions of U_{it} can differ across time periods. Assumption ID(ii) imposes conditional independence between α_i and U_{it} , which is crucial for identification. The assumption of conditional independence between U_{it} and $U_{i(-t)}$ can be relaxed; Remark 7 below and Section 6.2 in the Appendix explain how to identify the model with serially correlated disturbances U_{it} . Assumption ID(iii) is standard. Assumption ID(iv) is technical and very mild. Characteristic functions of most standard distributions do not vanish on the real line. For instance, Assumption ID(iv) is satisfied when, conditional on X_{it} , U_{it} has normal, log-normal, Cauchy, Laplace, χ^2 , or Student-t distribution. Assumption ID(v) is mild and guarantees the existence of conditional characteristic functions of $m(x, \alpha_i)$ and U_{it} . Assumption ID(vi) is restrictive, but is a key to identification. Inclusion of explanatory variables that violate this assumption (such as time-period dummies) into the model is discussed below in Remark 5. Assumption ID(vii) is standard. Note that Assumption ID(vii) is sufficient to identify the random effects model, but needs to be strengthened to identify the structural function when the effects α_i can be correlated with the covariates X_{it} . Assumption ID(viii) is not restrictive since function the $m(x, \alpha)$ can be a step function in α . Assumption ID(ix) is only needed when the covariates X_{it} contain continuously distributed components and is used to handle conditioning on certain probability zero events below. Finally, if the conditions of Assumption ID hold only for some, but not all, points of support of X_{it} let \mathcal{X} be the set of such points; then the identification results hold for all $x \in \mathcal{X}$.

For the clarity of exposition, identification of the random effects model is considered first. Subsequently, the result for the fixed effects is presented. Note that the random effect specification of model (1) can be interpreted as quantile regression with measurement error (U_{it}).¹⁴

The following assumption is standard in nonlinear random effect models:

Assumption RE. (i) α_i and X_i are independent; (ii) α_i has a uniform distribution on $[0, 1]$.

Assumption RE(i) defines the random effect model, while Assumption RE(ii) is a standard normalization, which is necessary since the function $m(x, \alpha)$ is modelled nonparametrically.

Theorem 1. *Suppose Assumptions ID and RE are satisfied. Then, model (1) is identified, i.e. functions $m(x, \alpha)$ and $f_{U_{it}|X_{it}}(u|x)$ are identified for all $x \in \mathcal{X}$, $\alpha \in (0, 1)$, $u \in \mathbb{R}$, and $t \in \{1, 2\}$.*

The proof of Theorem 1 uses the following extension of a result due to Kotlarski (1967) .

Lemma 1. *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) $E[U_1] = 0$,*

¹⁴I thank Victor Chernozhukov for suggesting this interpretation.

(iv) $\phi_{U_t}(s) \neq 0$ for all s and $t \in \{1, 2\}$. Then, the distributions of A , U_1 , and U_2 are identified from the joint distribution of (Y_1, Y_2) .

The proof of the lemma is given in the Appendix. See also Remark 9 below for the discussion.

Proof of Theorem 1. 1. Observe that $m(X_{i1}, \alpha_i) = m(X_{i2}, \alpha_i)$ when $X_{i1} = X_{i2} = x$. For any $x \in \mathcal{X}$,

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \Big| \{X_{i1} = X_{i2} = x\} = \begin{pmatrix} m(x, \alpha_i) + U_{i1} \\ m(x, \alpha_i) + U_{i2} \end{pmatrix} \Big| \{X_{i1} = X_{i2} = x\}. \quad (3)$$

Assumptions ID(i)-(v) ensure that Lemma 1 applies to (3), conditional on the event $X_{i1} = X_{i2} = x$, and identifies the conditional distributions of $m(x, \alpha_i)$, U_{i1} , and U_{i2} , given $X_{i1} = X_{i2} = x$, for all $x \in \mathcal{X}$. The conditional independence Assumption ID(ii) gives $f_{U_{it}|X_{i1}, X_{i2}}(u|x, x) = f_{U_{it}|X_{it}}(u|x)$ for $t \in \{1, 2\}$. That is, the conditional density $f_{U_{it}|X_{it}}(u|x)$ is identified for all $x \in \mathcal{X}$, $u \in \mathbb{R}$, and $t \in \{1, 2\}$.

2. Note that, conditional on $X_{it} = x$ (as opposed to $X_{i1} = X_{i2} = x$), the distribution of Y_{it} is a convolution of distributions of $m(x, \alpha_i)$ and U_{it} .¹⁵ Therefore, the conditional characteristic function of Y_{it} can be written as

$$\phi_{Y_{it}}(s|X_{it} = x) = \phi_{m(x, \alpha_i)}(s|X_{it} = x) \phi_{U_{it}}(s|X_{it} = x).$$

In this equation $\phi_{Y_{it}}(s|X_{it} = x) = E[\exp(isY_{it})|X_{it} = x]$ can be identified directly from data, while $\phi_{U_{it}}(s|X_{it} = x)$ was identified in the previous step. Thus, the conditional characteristic function of $m(x, \alpha_i)$ is identified by

$$\phi_{m(x, \alpha_i)}(s|X_{it} = x) = \frac{\phi_{Y_{it}}(s|X_{it} = x)}{\phi_{U_{it}}(s|X_{it} = x)},$$

where $\phi_{U_{it}}(s|X_{it} = x) \neq 0$ for all $s \in \mathbb{R}$, $x \in \mathcal{X}$, and $t \in \{1, 2\}$ due to Assumption ID(vi).

3. The identification of the characteristic function $\phi_{m(x, \alpha_i)}(s|X_{it} = x)$ is well known to be equivalent to the identification of the corresponding conditional distribution of $m(x, \alpha_i)$ given $X_{it} = x$; see, for example, Billingsley (1986, p. 355). For instance, the conditional cumulative distribution function $F_{m(x, \alpha_i)|X_{it}}(w|x)$ can be obtained using the result of Gil-Pelaez (1951):

$$F_{m(x, \alpha_i)|X_{it}}(w|x) = \frac{1}{2} - \lim_{\chi \rightarrow \infty} \int_{-\chi}^{\chi} \frac{e^{-isw}}{2\pi is} \phi_{m(x, \alpha_i)}(s|X_{it} = x) ds, \quad (4)$$

for any point $(w, x) \in (\mathbb{R}, \mathcal{X})$ of continuity of the CDF in w . Then, the conditional quantile function $Q_{m(x, \alpha_i)|X_{it}}(q|x)$ is identified from $F_{m(x, \alpha_i)|X_{it}}(w|x)$.

¹⁵A random variable Y is a convolution of independent random variables A and U if $Y = A + U$.

Finally, the structural function is identified for all $x \in \mathcal{X}$ and $\alpha \in (0, 1)$ by noticing that

$$Q_{m(x, \alpha_i) | X_{it}}(\alpha | x) = m(x, Q_{\alpha_i | X_{it}}(\alpha | x)) = m(x, \alpha), \quad (5)$$

where the first equality follows by the property of quantiles and Assumptions ID(vii), while the second equality follows from the normalization Assumption RE(ii).¹⁶17 ■

To gain some intuition for the first step of the proof, consider the special case when U_{i1} and U_{i2} are identically and symmetrically distributed, conditional on $X_{i2} = X_{i1} = x$ for all $x \in \mathcal{X}$. Then, for any $x \in \mathcal{X}$, the conditional characteristic function of $Y_{i2} - Y_{i1}$ equals

$$\begin{aligned} \phi_{Y_{i2} - Y_{i1}}(s | X_{i2} = X_{i1} = x) &= E[\exp\{is(Y_{i2} - Y_{i1})\} | X_{i2} = X_{i1} = x] \\ &= E[\exp\{is(U_{i2} - U_{i1})\} | X_{i2} = X_{i1} = x] \\ &= \phi_U(s|x) \phi_U(-s|x) = \phi_U(s|x)^2, \end{aligned} \quad (6)$$

where $\phi_U(s|x)$ denotes the characteristic function of U_{it} , conditional on $X_{it} = x$. Since U_{it} has a symmetric conditional distribution, $\phi_U(s|x)$ is symmetric in s and real valued. Thus, $\phi_U(s|x)$ is identified, because $\phi_{Y_{i2} - Y_{i1}}(s | X_{i2} = X_{i1} = x)$ is identified from data. Once $\phi_U(s|x)$ is known, the structural function $m(x, \alpha)$ is identified by following steps 2 and 3 of the proof.

Now consider the fixed effects model, i.e. the model where the unobserved heterogeneity α_i and covariates X_{it} can be correlated. A support assumption is needed to identify the fixed effects model. For any event ϑ , define $S_{\alpha_i}\{\vartheta\}$ to be the support of α_i , conditional on ϑ .

Assumption FE. (i) $m(x, \alpha)$ is strictly increasing in α ; (ii) for some $\bar{x} \in \mathcal{X}$ the normalization $m(\bar{x}, \alpha) = \alpha$ for all α is imposed; (iii) $S_{\alpha_i}\{(X_{it}, X_{i(-t)}) = (x, \bar{x})\} = S_{\alpha_i}\{X_{it} = x\}$ for all $x \in \mathcal{X}$ and $t \in \{1, 2\}$.

Assumption FE(i) is standard in the analysis of nonparametric models with endogeneity and guarantees invertibility of function $m(x, \alpha)$ in the second argument. Assumption FE(ii) is a mere normalization given Assumptions ID(viii) and FE(i). Assumption FE(iii) requires that the "extra" conditioning on $X_{it} = \bar{x}$ does not reduce the support of α_i . A conceptually similar support assumption is made by Altonji and Matzkin (2005). Importantly, neither their exchangeability assumption nor the index function assumption of Bester and Hansen (2007) are needed.

¹⁶In fact, the proof may be obtained in a shorter way. The first step identifies the conditional distribution of $m(x, \alpha_i)$, given the event $X_{i1} = X_{i2} = x$, and hence identifies $Q_{m(x, \alpha_i) | X_{i1}, X_{i2}}(q|x, x)$. Then, similar to (5) one obtains

$$Q_{m(x, \alpha_i) | X_{i1}, X_{i2}}(q|x, x) = m(x, Q_{\alpha_i | X_{i1}, X_{i2}}(\alpha|x, x)) = m(x, \alpha).$$

The proof of the theorem is presented in the three steps to keep it parallel to the proof of identification of the correlated random effects model in the next theorem.

¹⁷When X_{it} contains continuously distributed components the proof uses conditioning on probability zero events. Section 6.3 in the Appendix formally establishes that this conditioning is valid under Assumption ID(ix).

Theorem 2. *Suppose Assumptions ID and FE(i)-(ii) hold. Then, in model (1) the structural function $m(x, \alpha)$ and the conditional and unconditional distributions of the unobserved heterogeneity $F_{\alpha_i}(\alpha|X_{it} = x)$ and $F_{\alpha_i}(\alpha)$, and the idiosyncratic disturbances $F_{U_{it}}(u|X_{it} = x)$ are identified for all $x \in \mathcal{X}$, $\alpha \in S_{\alpha_i} \{(X_{it}, X_{i\tau}) = (x, \bar{x})\}$, $u \in \mathbb{R}$, and $t \in \{1, 2\}$.*

If in addition Assumption FE(iii) is satisfied, the functions $m(x, \alpha)$ and $F_{\alpha_i}(\alpha|X_{it} = x)$ are identified for all $x \in \mathcal{X}$, $\alpha \in S_{\alpha_i} \{X_{it} = x\}$, and $t \in \{1, 2\}$.

Proof. 1. Identify the conditional distribution of U_{it} given X_{it} exactly following the first step of the proof of Theorem 1. In particular, the conditional characteristic functions $\phi_{U_{it}}(s|X_{it} = x)$ are identified for all $t \in \{1, 2\}$.

2. Take any $x \in \mathcal{X}$ and note that the conditional characteristic functions of Y_{i1} and Y_{i2} , given the event $(X_{i1}, X_{i2}) = (x, \bar{x})$ satisfy

$$\begin{aligned}\phi_{Y_{i1}}(s|X_{i1} = x, X_{i2} = \bar{x}) &= \phi_{m(X_{i1}, \alpha_i)}(s|X_{i1} = x, X_{i2} = \bar{x}) \phi_{U_{i1}}(s|X_{i1} = x), \\ \phi_{Y_{i2}}(s|X_{i1} = x, X_{i2} = \bar{x}) &= \phi_{\alpha_i}(s|X_{i1} = x, X_{i2} = \bar{x}) \phi_{U_{i2}}(s|X_{i2} = \bar{x}),\end{aligned}$$

where Assumption FE(ii) is used in the second line. The conditional characteristic functions of Y_{it} on the left-hand sides of the equations are identified from data, while the function $\phi_{U_{it}}(s|X_{it} = x)$ is identified in step 1 of the proof. Due to Assumption ID(iv), $\phi_{U_{it}}(s|x) \neq 0$ for all $s \in \mathbb{R}$, $x \in \mathcal{X}$, and $t \in \{1, 2\}$. Thus, we can identify the following conditional characteristic functions:

$$\phi_{m(x, \alpha_i)}(s|X_{i1} = x, X_{i2} = \bar{x}) = \frac{\phi_{Y_{i1}}(s|X_{i1} = x, X_{i2} = \bar{x})}{\phi_{U_{i1}}(s|X_{i1} = \bar{x})}, \quad (7)$$

$$\phi_{\alpha_i}(s|X_{i1} = x, X_{i2} = \bar{x}) = \frac{\phi_{Y_{i2}}(s|X_{i1} = x, X_{i2} = \bar{x})}{\phi_{U_{i2}}(s|X_{i2} = \bar{x})}. \quad (8)$$

As explained in the proof of Theorem 1, these conditional characteristic functions uniquely determine the quantile function $Q_{m(x, \alpha_i)}(q|X_{i1} = x, X_{i2} = \bar{x})$ and the cumulative distribution function $F_{\alpha_i|X_{i1}, X_{i2}}(a|x, \bar{x})$.

3. Then, the structural function $m(x, \alpha)$ is identified by

$$\begin{aligned}& Q_{m(x, \alpha_i)|X_{i1}, X_{i2}}(F_{\alpha_i|X_{i1}, X_{i2}}(a|x, \bar{x})|x, \bar{x}) \\ &= m(x, Q_{\alpha_i|X_{i1}, X_{i2}}(F_{\alpha_i|X_{i1}, X_{i2}}(a|x, \bar{x})|x, \bar{x})) = m(x, a),\end{aligned} \quad (9)$$

where the first equality follows by the property of quantiles and Assumption FE(i) and the second equality follows from the definition of the quantile function and Assumption ID(viii).

Next, consider identification of $F_{\alpha_i}(\alpha|X_{it} = x)$. Similar to step 2, function $\phi_{Y_{it}}(s|X_{it} = x)$

is identified from data and hence

$$\phi_{m(X_{it}, \alpha_i)}(s|X_{it} = x) = \phi_{Y_{it}}(s|X_{it} = x) / \phi_{U_{it}}(s|X_{it} = x) \quad (10)$$

is identified. Hence, the quantile function $Q_{m(x, \alpha_i)|X_{it}}(q|x)$ is identified for all $x \in \mathcal{X}$, $q \in (0, 1)$, and $t \in \{1, 2\}$. Then, by the property of quantiles

$$Q_{m(X_{it}, \alpha_i)|X_{it}}(q|x) = m(x, Q_{\alpha_i|X_{it}}(q|x)) \text{ for all } q \in (0, 1).$$

Thus, using Assumption FE(i) the conditional distribution of α_i is identified by

$$Q_{\alpha_i|X_{it}}(q|x) = m^{-1}(x, Q_{m(X_{it}, \alpha_i)|X_{it}}(q|x)), \quad (11)$$

where $m^{-1}(x, w)$ denotes the inverse of $m(x, \alpha)$ in the second argument, which is identified for all $\alpha \in S_{\alpha_i}\{X_{it} = x\}$ when Assumption FE(iii) holds. Finally, one identifies the conditional cumulative distribution function $F_{\alpha_i|X_{it}}(a|x)$ by inverting the quantile function $Q_{\alpha_i|X_{it}}(q|x)$. Then, the unconditional cumulative distribution function is identified by

$$\int F_{\alpha_i}(a|X_{it} = x) f_{X_{it}}(x) dx,$$

where the conditional density $f_{X_{it}}(x)$ is identified directly from data. ■

Remark 1. *Function $m(x, \alpha)$ and the distribution of α_i depend on the choice of \bar{x} . However, it is easy to show that function*

$$h(x, q) \equiv m(x, Q_{\alpha_i}(q))$$

does not depend on the choice of normalization \bar{x} . The function $h(x, q)$ is of interest for policy analysis. In the union membership example, the value $h(1, 0.5) - h(0, 0.5)$ is the union membership premium for a person with median skill level.

Remark 2. *Assumption FE(iii) may fail in some applications. In this case Theorem 2 secures point identification of function $m(x, \alpha)$ for all $\alpha \in S_{\alpha_i}\{(X_{i1}, X_{i2}) = (x, \bar{x})\}$, while the set $S_{\alpha_i}\{(X_{i1}, X_{i2}) = (x, \bar{x})\}$ is a strict subset of the set $S_{\alpha_i}\{X_{i1} = x\}$. However, it is likely that $[Q_{\alpha_i}(\underline{q}|X_{i1} = x), Q_{\alpha_i}(\bar{q}|X_{i1} = x)] \subset S_{\alpha_i}\{(X_{i1}, X_{i2}) = (x, \bar{x})\}$ for some small value \underline{q} and a large value \bar{q} such that $0 \leq \underline{q} < \bar{q} \leq 1$. In other words, it is likely that even when Assumption FE(iii) fails, one obtains identification of the function $m(x, \alpha)$ (and correspondingly $h(x, q)$) for most values of α (and q) except the most extreme ones.¹⁸*

¹⁸Consider the union wage premium example. Suppose that individuals with unobserved skill level α never join the union (for instance, because α is too low). Then $\alpha \in S_{\alpha_i}\{X_{i1} = 0\}$, but $\alpha \notin S_{\alpha_i}\{(X_{i1}, X_{i2}) = (0, 1)\}$. In this case, failure of identification is the property of the economic data and not of the model. The population

Remark 3. *The third steps of Theorems 1 and 2 can be seen as the distributional generalizations of between- and within-variation analysis, respectively. The works of Altonji and Matzkin (2005) and Athey and Imbens (2006) use distributional manipulations similar to the third step of Theorem 2. However, their models contain only a scalar unobservable and hence the quantile transformations apply directly to the observed distributions of outcomes. The method of this paper instead filters out the idiosyncratic disturbances at the first stage, and only after that uses between- or within-variation for identification.*

Remark 4. *Time effects can be added into the model, i.e. the model*

$$Y_{it} = m(X_{it}, \alpha_i) + \eta_t(X_{it}) + U_{it} \quad (12)$$

is identified. Indeed, normalize $\eta_1(x) = 0$ for all $x \in \mathcal{X}$, then for any $t > 1$ time effects $\eta_t(x)$ are identified from

$$E[Y_{it} - Y_{i1} | X_{it} = X_{i1} = x] = E[U_{it} + \eta_t(X_{it}) - U_{i1} | X_{it} = X_{i1} = x] = \eta_t(x).$$

Once the time effects are identified, identification of the rest of the model proceeds as described above, except the random variable Y_{it} is replaced by $Y_{it} - \eta_t(X_{it})$.

Remark 5. *The identification strategies of Theorems 1 and 2 require that the joint density of (X_{i1}, X_{i2}) is positive at (x, x) , i.e. $f_{X_{i1}, X_{i2}}(x, x) > 0$, see Assumption ID(vii). In some situations this may become a problem, for example, if an individual's age is among the explanatory variables. Suppose that the joint density $f_{(Z_{it}, Z_{i\tau})}(z, z) = 0$, where Z_{it} are some explanatory variables, different from the elements of X_{it} . Assume that α_i and Z_i are independent, conditional on X_i , and $E[U_i | X_i, Z_i] = 0$. The following model is identified:*

$$Y_{it} = m(X_{it}, \alpha_i) + g(X_{it}, Z_{it}) + U_{it}.$$

Note that the time effects $\eta_t(x)$ of the model (12) are a special case of $g(x, z)$ with $Z_{it} = t$. To separate the functions $m(\cdot)$ and $g(\cdot)$, impose the normalization $g(x, z_0) \equiv 0$ for some point z_0 and all $x \in \mathcal{X}$. Then, identification follows from the fact that

$$g(x, z) = E[Y_{it} - Y_{i\tau} | X_{it} = X_{i\tau} = x, Z_{it} = z, Z_{i\tau} = z_0],$$

under the assumption that the conditional expectation of interest is observable. Then, define $\tilde{Y}_{it} = Y_{it} - g(X_{it}, Z_{it})$ and proceed as before with \tilde{Y}_{it} instead of Y_{it} .

contains no individuals with skill level α holding union jobs, and hence there is no way of identifying the union wage for people with such skill level (at least without imposing strong assumptions permitting extrapolation, such as functional form assumptions).

Remark 6. *In some applications the support of the covariates in the first time period X_{i1} is smaller than the support of covariates in the second period. For instance, suppose X_{it} is the treatment status (0 or 1) of an individual i in period t and no one is treated in the first period, but some people are treated in the second period. Then, the event $X_{i1} = X_{i2} = 0$ has positive probability, but there are no individuals with $X_{i1} = X_{i2} = 1$ in the population, thus Assumption ID(vi) is violated. A solution is to assume that the idiosyncratic disturbances U_{it} are independent of the treatment status X_{it} . Then, the distribution of the disturbances U_{it} is identified by conditioning on the event $X_{i1} = X_{i2} = 0$ and the observations with $X_{i1} = X_{i2} = 1$ are not needed. Identification of the structural function $m(x, \alpha)$ only requires conditioning on the event $\{X_{i0} = 0, X_{i1} = 1\}$, which has positive probability.*

Remark 7. *Assumption ID(ii) does not permit serial correlation of the disturbances. It is possible to relax this assumption. Section 6.2 in the Appendix shows how to identify model (1) when the disturbance U_{it} follows AR(1) or MA(1) processes. The identification argument requires panel data with three time periods. It is important to note that the assumption of conditional independence between the unobserved heterogeneity α_i and the disturbances U_{it} is still a key to identification. However, this assumption requires some further restrictions regarding the initial disturbance U_{i1} . As discussed in Section 6.2, in panel models with serially correlated disturbances, the initial disturbance U_{i1} and the unobserved heterogeneity α_i can be dependent when the past innovations are conditionally heteroskedastic. On the one hand, due to conditional heteroskedasticity, the initial disturbance U_{i1} depends on the past innovations and hence on the past (unobserved) covariates. On the other hand, the unobserved heterogeneity α_i may also correlate with the past covariates and hence with the initial disturbance U_{i1} . Section 6.2 provides assumptions that are sufficient to guarantee the conditional independence of α_i and U_{i1} .*

Remark 8. *The results of this section can be extended to the case of misclassified discrete covariates if the probability of misclassification is known or can be identified from a validation dataset. Such empirical settings have for example been considered by Card (1996). When the probability of misclassification is known, it is possible to back out conditional characteristic functions of (Y_{i1}, Y_{i2}) given the true values of covariates $(\phi_{Y_{i1}, Y_{i2}}(s_1, s_2 | X_{i1}^*, X_{i2}^*))$ from the conditional characteristic functions of (Y_{i1}, Y_{i2}) given the misclassified values of covariates $(\phi_{Y_{i1}, Y_{i2}}(s_1, s_2 | X_{i1}, X_{i2}))$. Once the characteristic functions $\phi_{Y_{i1}, Y_{i2}}(s_1, s_2 | X_{i1}^*, X_{i2}^*)$ are identified the identification of function $m(x, \alpha)$ proceeds as in the proofs of Theorems 1 and 2. The assumptions and the details of identification strategy are presented in Section 6.4 in the Appendix.*

Remark 9. *This paper aims to impose minimal assumptions on function $m(x, \alpha)$ and the conditional distribution of $m(x, \alpha_i)$. The original result of Kotlarski (1967) requires the assumption that the characteristic function $\phi_{m(x, \alpha_i)}(\cdot | X_{it} = x)$ is nonvanishing. This assump-*

tion may be violated, for instance, when $m(x, \alpha_i)$ has a discrete distribution. Lemma 1 avoids imposing this assumption on the distribution of $m(x, \alpha_i)$.

3 Estimation

The proofs of Theorems 1 and 2 are constructive and hence suggest a natural way of estimating the quantities of interest. Conditional characteristic functions are the building blocks of the identification strategy. In Section 3.1 they are replaced by their empirical analogs to construct estimators. Essentially, the estimation method requires performing deconvolution conditional on the values of the covariates.

When the covariates are *discrete*, estimation can be performed using the existing deconvolution techniques. The sample should be split into subgroups according to the values of the covariates and a deconvolution procedure should be used to obtain the estimates of necessary cumulative distribution functions, see the expressions for the estimators $\hat{m}_{RE}(x, \alpha)$ and $\hat{m}_{FE}(x, \alpha)$ below. There is a large number of deconvolution techniques in statistics literature that can be used in this case. For example, the kernel deconvolution method is well studied, see the recent papers of Delaigle, Hall, and Meister (2008) and Hall and Lahiri (2008) for the description of this method as well as for a list of alternative approaches.¹⁹

In contrast, when the covariates are *continuously distributed* one needs to use a *conditional deconvolution* procedure. To the best of my knowledge this is the first paper to propose conditional deconvolution estimators as well as to study their statistical properties.

Section 3.1 presents estimators of conditional the cumulative distribution functions and the conditional quantile functions by means of conditional deconvolution. Section 3.1 also provides estimators of the structural function $m(x, \alpha)$ and the conditional distribution of heterogeneity α_i . Importantly, the estimators of the structural functions are given by an explicit formula and require no optimization. Section 3.2 derives the rates of convergence for the proposed estimators. As a by-product of this derivation, the section obtains the rates of convergence of the proposed conditional cumulative distribution and quantile function estimators in the problem of conditional deconvolution.

In this section T can be 2 or larger. All limits are taken as $n \rightarrow \infty$ for T fixed. For simplicity, the panel dataset is assumed to be balanced. To simplify the notation below, $x_{t\tau}$ stands for (x_t, x_τ) . In addition, the conditioning notation " $|x_t$ " and " $|x_{t\tau}$ " should be, correspondingly, read as "conditional on $X_{it} = x_t$ " and "conditional on $(X_{it}, X_{i\tau}) = (x_t, x_\tau)$ ". Thus, for example, $F_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ means $F_{m(\alpha_i, x_t)|X_{it}, X_{i\tau}}(\omega|x_t, x_\tau)$.

¹⁹The number of econometric applications of deconvolution methods in econometrics is small but growing; an incomplete list includes Horowitz and Markatou (1996), Li and Vuong (1998), Heckman, Smith, and Clements (1997), Schennach (2004a), Hu and Ridder (2005), and Bonhomme and Robin (2008).

3.1 Conditional CDF and Quantile Function Estimators

Remember that the conditional characteristic function $\phi_{Y_{it}}(s|x_t)$ is simply the conditional expectation; $\phi_{Y_{it}}(s|x_t) = E[\exp(isY_{it})|X_{it} = x_t]$. Therefore, it is natural to estimate it by the Nadaraya-Watson kernel estimator²⁰

$$\hat{\phi}_{Y_{it}}(s|x_t) = \frac{\sum_{i=1}^n \exp(isY_{it}) K_{h_Y}(X_{it} - x_t)}{\sum_{i=1}^n K_{h_Y}(X_{it} - x_t)},$$

where $h_Y \rightarrow 0$ is a bandwidth parameter, $K_h(\cdot) \equiv K(\cdot/h)/h$ and $K(\cdot)$ is a standard kernel function.²¹

When the conditional distribution of U_{it} is assumed to be symmetric and the same across t , formula (6) identifies its characteristic function. In this case $\phi_U(s|x) = \phi_{U_{it}}(s|X_{it} = x)$ is the same for all t and can be estimated by

$$\hat{\phi}_U^S(s|x) = \left| \frac{\sum_{t=1}^{T-1} \sum_{\tau=t+1}^T \sum_{i=1}^n \exp\{is(Y_{it} - Y_{i\tau})\} K_{h_U}(X_{it} - x) K_{h_U}(X_{i\tau} - x)}{\sum_{t=1}^{T-1} \sum_{\tau=t+1}^T \sum_{i=1}^n K_{h_U}(X_{it} - x) K_{h_U}(X_{i\tau} - x)} \right|^{1/2},$$

where bandwidth $h_U \rightarrow 0$. If the researcher is unwilling to impose the assumptions of symmetry and distributional equality across the time periods, an alternative estimator of $\phi_{U_{it}}(s|x_t)$ can be based on the empirical analog of equation (A.1):

$$\hat{\phi}_{U_{it}}^{AS}(s|x_t) = \frac{1}{T-1} \sum_{\substack{\tau=1 \\ \tau \neq t}}^T \exp \left\{ i \int_0^s \frac{\sum_{i=1}^n Y_{it} e^{i\xi(Y_{it} - Y_{i\tau})} K_{h_U}(X_{it} - x_t) K_{h_U}(X_{i\tau} - x_t)}{\sum_{i=1}^n e^{i\xi(Y_{it} - Y_{i\tau})} K_{h_U}(X_{it} - x_t) K_{h_U}(X_{i\tau} - x_t)} d\xi - is \frac{\sum_{i=1}^n Y_{it} K_{h_Y}(X_{it} - x_t)}{\sum_{i=1}^n K_{h_Y}(X_{it} - x_t)} \right\}.$$

In what follows an estimator of $\phi_{U_{it}}(s|x_t)$ is written as $\hat{\phi}_{U_{it}}(s|x_t)$ and means either $\hat{\phi}_{U_{it}}^S(s|x_t)$ or $\hat{\phi}_{U_{it}}^{AS}(s|x_t)$ depending on the assumptions that the researcher imposes about U_{it} .

Using (10), the conditional characteristic function $\phi_{m(x_t, \alpha_i)}(s|x_t)$ can be estimated by $\hat{\phi}_{m(x_t, \alpha_i)}(s|x_t) = \hat{\phi}_{Y_{it}}(s|x_t) / \hat{\phi}_{U_{it}}(s|x_t)$. Then, the goal is to estimate the conditional cumulative distribution functions $F_{m(x_t, \alpha_i)}(\omega|x_t)$ and the corresponding quantile functions $Q_{m(x_t, \alpha_i)}(q|x_t)$

²⁰In this section all covariates are assumed to be continuous. As usual, estimation in the presence of discrete covariates can be performed by splitting the sample into subsamples, according to the values of the discrete covariates. Naturally, the number of discrete covariates does not affect the rates of convergence. When the model does not contain any continuous covariates the standard (unconditional) deconvolution procedures can be used (see for example Fan, 1991).

²¹As usual, the data may need to be transformed so that all the elements of the vector of covariates X_{it} have the same magnitude.

for each t . The estimator of $F_{m(x_t, \alpha_i)}(\omega|x_t)$ can be based on equation (4):

$$\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t) = \frac{1}{2} - \int_{-\infty}^{\infty} \frac{e^{-is\omega}}{2\pi is} \phi_w(h_w s) \frac{\hat{\phi}_{Y_{it}}(s|x_t)}{\hat{\phi}_{U_{it}}(s|x_t)} ds, \quad (13)$$

where $\phi_w(\cdot)$ is the Fourier transform of a kernel function $w(\cdot)$ and $h_w \rightarrow 0$ is a bandwidth parameter. The kernel function $w(\cdot)$ should be such that its Fourier transform $\phi_w(s)$ has bounded support. For example, the so-called sinc kernel $w_{\text{sinc}}(s) = \sin(s)/(\pi s)$ has Fourier transform $\phi_{w_{\text{sinc}}}(s) = I(|s| \leq 1)$. More details on $\phi_w(\cdot)$ are given below. Use of the smoothing function $\phi_w(\cdot)$ is standard in the deconvolution literature, see for example Fan (1991) and Stefanski and Carroll (1990). The smoothing function $\phi_w(\cdot)$ is necessary because deconvolution is an ill-posed inverse problem. The ill-posedness manifests itself in the poor behavior of the ratio $\hat{\phi}_{Y_{it}}(s|x_t)/\hat{\phi}_{U_{it}}(s|x_t)$ for large $|s|$, since both $\phi_{U_{it}}(s|x_t)$ and its consistent estimator $\hat{\phi}_{U_{it}}(s|x_t)$ approach zero as $|s| \rightarrow \infty$.

The estimator of $F_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ is similar to (13). Define

$$\hat{\phi}_{Y_{it}}(s|x_{t\tau}) = \frac{\sum_{i=1}^n \exp\{isY_{it}\} K_{h_Y}(X_{it} - x_t) K_{h_Y}(X_{i\tau} - x_{t\tau})}{\sum_{i=1}^n K_{h_Y}(X_{it} - x_t) K_{h_Y}(X_{i\tau} - x_{t\tau})}.$$

Then, the estimator $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ is exactly the same as $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ except $\hat{\phi}_{Y_{it}}(s|x_t)$ is replaced by $\hat{\phi}_{Y_{it}}(s|x_{t\tau})$. Note that the choice of bandwidth h_Y is different for $\hat{\phi}_{Y_{it}}(s|x_t)$ and $\hat{\phi}_{Y_{it}}(s|x_{t\tau})$. Below, $h_Y(d)$ stands for the bandwidth used for estimation of $\hat{\phi}_{Y_{it}}(s|x_t)$ and $\hat{\phi}_{Y_{it}}(s|x_{t\tau})$, respectively, when $d = p$ and $d = 2p$, where p is the number of covariates, i.e. the length of vector X_{it} .

Conditional quantiles of the distribution of $m(x_t, \alpha_i)$ can be estimated by the inverse of the corresponding conditional CDFs. The estimates of CDFs can be non-monotonic, thus a monotone version of CDFs should be inverted. This paper uses the rearrangement technique proposed by Chernozhukov, Fernandez-Val, and Galichon (2007). Function $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ is estimated on a fine grid²² of values ω and the estimated values of $\hat{F}_{m(x_t, \alpha_i)}$ are then sorted in increasing order. The resulting CDF is monotone increasing by construction. Moreover, Chernozhukov, Fernandez-Val, and Galichon (2007) show that this procedure improves the estimates of CDFs and quantile functions in finite samples. Define $\tilde{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ to be the rearranged version of $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$.

²²In practice, it is suggested to take a fine grid is taken on the interval $[\hat{Q}_{Y_{it}|X_{it}}(\delta|x_t), \hat{Q}_{Y_{it}|X_{it}}(1 - \delta|x_t)]$, where $\hat{Q}_{Y_{it}|X_{it}}(q|x_t)$ is the estimator of $Q_{Y_{it}|X_{it}}(q|x_t)$, i.e. of the q -th conditional quantile of Y_{it} given $X_{it} = x_t$, and δ is a very small number, such as 0.01. The conditional quantiles $Q_{Y_{it}|X_{it}}(q|x_t)$ can be estimated by the usual kernel methods, see for example Bhattacharya and Gangopadhyay (1990).

The conditional quantile function $Q_m(q|x_t)$ can then be estimated by

$$\hat{Q}_{m(x_t, \alpha_i)}(q|x_t) = \min_{\omega} \left\{ \tilde{F}_{m(x_t, \alpha_i)}(\omega|x_t) \geq q \right\},$$

where ω takes values on the above-mentioned grid. Estimation of $Q_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ is analogous to estimation of $Q_{m(x_t, \alpha_i)}(\omega|x_t)$.

As always, it is hard to estimate quantiles of a distribution in the areas where the density is close to zero. In particular, it is hard to estimate the quantiles in the tails of a distribution, i.e. the so-called extreme quantiles. Therefore, it is suggested to restrict attention only to the estimates of, say, 0.05-th to the 0.95-th quantiles of the distribution of $m(x_t, \alpha_i)$.²³

Once conditional CDFs and Quantile Functions are estimated, formula (5) suggests the following estimator of the structural function $m(x, \alpha)$ in the random effects model:

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

Note that compared to expression (5) the estimator \hat{m}_{RE} contains extra averaging across time periods to improve finite sample performance. As usual x should take values from the interior of set \mathcal{X} to avoid kernel estimation on the boundary of the set \mathcal{X} . As mentioned earlier, one should only consider "not too extreme" values of α , say $\alpha \in [0.05, 0.95]$. At the same time, one does not have to estimate the conditional distribution of α_i since in the random effect model $Q_{\alpha_i}(q|x_t) \equiv q$ for all $q \in (0, 1)$ and $x_t \in \mathcal{X}$ by Assumption RE(ii).

In the fixed effects model it is possible to use the empirical analog of equation (9) to estimate $m(x, \alpha)$. Instead, a better procedure is suggested based on the following formula that (also) identifies the function $m(x, \alpha)$. Following the arguments of the proof of Theorem 2 it is easy to show that for all $x \in \mathcal{X}$, $x_2 \in \mathcal{X}$, t , and $\tau \neq t$, the structural function can be written as

$$m(x, a) = Q_{m(x, \alpha_i)|X_{i,t\tau}} \left(F_{m(x_2, \alpha_i)|X_{i,t\tau}} \left(Q_{m(x_2, \alpha_i)|X_{i,t\tau}} \left(F_{m(\bar{x}, \alpha_i)|X_{i,t\tau}} \left(a|\bar{x}, x_2 \right) | \bar{x}, x_2 \right) | x, x_2 \right) | x, x_2 \right),$$

where $X_{i,t\tau}$ stands for $(X_{it}, X_{i\tau})$ to shorten the notation. Note that x_2 in the above formula only enters on the right-hand side but not on the left-hand side of the equation. Thus, the

²³Of course, confidence intervals should be used for inference. However, derivation of confidence intervals for extreme quantiles is itself a hard problem and is not considered in this paper.

estimator proposed below averages over x_2 :

$$\hat{m}_{FE}(x, a) = \frac{1}{T(T-1)} \sum_{t=1}^T \sum_{\tau=1, \tau \neq t}^T \frac{1}{leb(\bar{\mathcal{X}})} \times \int_{\bar{\mathcal{X}}} \hat{Q}_{m(x, \alpha_i) | X_{i,t\tau}}(\tilde{F}_{m(x_2, \alpha_i) | X_{i,t\tau}}(\hat{Q}_{m(x_2, \alpha_i) | X_{i,t\tau}}(\tilde{F}_{m(\bar{x}, \alpha_i) | X_{i,t\tau}}(a | \bar{x}, x_2) | \bar{x}, x_2) | x, x_2) | x, x_2) dx_2, \quad (14)$$

for all $x \in \bar{\mathcal{X}}$, where set $\bar{\mathcal{X}}$ is a subset of \mathcal{X} that does not include the boundary points of \mathcal{X} , and the normalization factor $leb(\bar{\mathcal{X}})$ is the Lebesgue measure of set $\bar{\mathcal{X}}$. Averaging over x_2 , t , and τ exploits the overidentifying restrictions that are present in the model. Once again, the values of a should not be taken to be "too extreme". The set $\bar{\mathcal{X}}$ is introduced to avoid estimation at the boundary points. In practice, integration over x_2 can be substituted by summation over a grid over $\bar{\mathcal{X}}$ with the cell size of order $h_Y(2p)$ or smaller. Then, instead of $leb(\bar{\mathcal{X}})$, the normalization factor should be the number of grid points.²⁴

Note that the ranges and domains of the estimators in the above equations are bounded intervals that may not be compatible. In other words, for some values of a and x_2 the expression $\hat{Q}\left(\tilde{F}\left(\hat{Q}\left(\tilde{F}(a | \bar{x}, x_2) | \bar{x}, x_2\right) | x, x_2\right) | x, x_2\right)$ cannot be evaluated because the value of one of the functions $\tilde{F}(\cdot)$ or $\hat{Q}(\cdot)$ at the corresponding argument is not determined (for some very large or very small a). This can happen because the supports of the corresponding conditional distributions may not be the same or, more likely, due to the finite sample variability of estimators. Moreover, consider fixing the value of a and varying x_2 . It can be that for some values of x_2 the argument of one of the estimated functions $\tilde{F}(\cdot)$ or $\hat{Q}(\cdot)$ corresponds to extreme quantiles, while for other values of x_2 the arguments of all estimated CDFs and quantile functions correspond to intermediate quantiles. It is suggested to drop the values of $\hat{Q}\left(\tilde{F}\left(\hat{Q}\left(\tilde{F}(a | \bar{x}, x_2) \dots\right)\right)\right)$ that rely on estimates at extreme quantiles from the averaging and adjust the normalization factor correspondingly.²⁵

Finally, using (11) the conditional quantiles of the distribution of α_i is estimated by

$$\hat{Q}_{\alpha_i}(q | X_{it} = x) = \hat{m}_{FE}^{-1}\left(x, \hat{Q}_{m(x, \alpha_i) | X_{it}}(q | x)\right).$$

If the conditional distribution is assumed to be stationary, i.e. if $Q_{\alpha_i}(q | X_{it} = x)$ is believed to be the same for all t then additional averaging over t can be used to obtain

$$\hat{Q}_{\alpha_i}(q | X_{it} = x) = \frac{1}{T} \sum_{\tau=1}^T \hat{m}_{FE}^{-1}\left(x, \hat{Q}_{m(x, \alpha_i) | X_{i\tau}}(q | x)\right).$$

²⁴When X_{it} is discrete one should substitute integration with the summation over the points of support of the distribution of X_{it} .

²⁵Note that \hat{m}_{FE} depends on the choice of \bar{x} . In practice, it is suggested to take the value of \bar{x} to be the mode of the distribution of X_{it} because the variance of the kernel estimators is inversely proportional to the density of X_{it} at a particular point.

Then, by inverting the conditional quantile functions one can use the sample analog of the formula $F_{\alpha_i}(\alpha) = \int_{\mathcal{X}} F_{\alpha_i}(\alpha|x_t) f(x_t) dx_t$ to obtain the unconditional CDF $\hat{F}_{\alpha_i}(a)$ and the quantile function $\hat{Q}_{\alpha_i}(q)$. Then, the policy relevant function $h(x, q)$, discussed in Remark 1, can be estimated by $\hat{h}(x, q) = \hat{m}_{FE}(x, \hat{Q}_{\alpha_i}(q))$.

3.2 Rates of Convergence of the Estimators

First, this section derives the rates of convergence of the conditional deconvolution estimators $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ and $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$. Then, it provides the rates of convergence for the other estimators proposed in the previous section.

Although the estimators of characteristic functions $\hat{\phi}_{Y_{it}}(\cdot)$ and $\hat{\phi}_{U_{it}}(\cdot)$ proposed in the previous section have typical kernel estimator form, several issues are to be tackled when deriving the rate of convergence of the estimators of the CDFs $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$. First, as discussed in the previous section, for any distribution of errors U_{it} the characteristic function $\phi_{U_{it}}(s|x)$ approaches zero for large $|s|$, i.e. $\phi_{U_{it}}(s|x) \rightarrow 0$ as $|s| \rightarrow \infty$. This creates a problem, since the estimator of $\phi_{U_{it}}(s|x)$ is in the denominator of the second fraction in (13). Moreover, the bias of the kernel estimator of the conditional expectation $E[\exp\{isY_{it}\} | X_{it} = x]$ grows with $|s|$. Thus, large values of $|s|$ require special treatment. In addition, extra care also needs to be taken when s is in the vicinity of zero because of the term $1/s$ in formula (13). For instance, obtaining the rate of convergence of suprema of $|\hat{\phi}_{Y_{it}}(s|x) - \phi_{Y_{it}}(s|x)|$ and $|\hat{\phi}_{U_{it}}(s|x) - \phi_{U_{it}}(s|x)|$ is insufficient to deduce the rate of convergence of the estimator $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$. Instead, the lemmas in the appendix derive the rates of convergence for the suprema of $|s^{-1}(\hat{\phi}_{Y_{it}}(s|x) - \phi_{Y_{it}}(s|x))|$ and $|s^{-1}(\hat{\phi}_{U_{it}}(s|x) - \phi_{U_{it}}(s|x))|$ over the expanding set $s \in [-h_w^{-1}, 0) \cup (0, h_w^{-1}]$ and all x . Then, these results are used to obtain the rate of convergence of the integral in the formula for $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$.

The assumptions made in this section reflect the difficulties specific to the problem of conditional deconvolution. The first three assumptions are, however, standard:

Assumption 1. $F_{m(x_t, \alpha_i)}(\omega|x_t)$ and $F_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ have $k_F \geq 1$ continuous absolutely integrable derivatives with respect to ω for all $\omega \in \mathbb{R}$, all $x_{t\tau} \in \mathcal{X} \times \mathcal{X}$, and all $t, \tau \neq t$.

For any set \mathcal{Z} , denote $\mathcal{D}_{\mathcal{Z}}(k)$ to be the class of functions $\varphi : \mathcal{Z} \rightarrow \mathbb{C}$ with k continuous mixed partial derivatives.

Assumption 2. (i) $\mathcal{X} \subset \mathbb{R}^p$ is bounded and the joint density $f_{(X_t, X_\tau)}(x_{t\tau})$ is bounded from above and is bounded away from zero for all $t, \tau \neq t$, and $x_{t\tau} \in \mathcal{X} \times \mathcal{X}$; (ii) there is a positive integer $\bar{k} \geq 1$ such that $f_{X_t}(x_t) \in \mathcal{D}_{\mathcal{X}}(\bar{k})$ and $f_{(X_t, X_\tau)}(x_{t\tau}) \in \mathcal{D}_{\mathcal{X} \times \mathcal{X}}(\bar{k})$.

Assumption 3. There is a constant $\gamma > 2$ such that $\sup_{x_{t\tau} \in \mathcal{X} \times \mathcal{X}} E[|Y_{it}|^\gamma | x_{t\tau}]$ is bounded for all t and τ .

Assumption 1 is only slightly stronger than the usual assumption of bounded derivatives. The integer \bar{k} introduced in Assumption 2(ii) is also used in Assumptions 4 and 5 below. The value \bar{k} should be taken so that these assumptions hold as well. It may be helpful to think of \bar{k} as of the "smoothness in x_t (or $x_{t\tau}$)".

The rate of convergence of deconvolution estimators depends critically on the rate of decrease of the characteristic function of the errors. Therefore, for all positive numbers \bar{s} it is useful to define the following function:

$$\chi(\bar{s}) = \max_{1 \leq t \leq T} \sup_{(s,x) \in [-\bar{s}, \bar{s}] \times \mathcal{X}} 1/|\phi_{U_{it}}(s|X_{it} = x)|.$$

It is a well-known result that $\chi(\bar{s}) \rightarrow \infty$ as $\bar{s} \rightarrow \infty$ for any distribution of U_{it} . Estimator (13) contains $1/\hat{\phi}_{U_{it}}(s|x_t)$ as a factor, hence the rate of convergence of the estimator will depend on the rate of growth of $\chi(\bar{s})$. The following two alternative assumptions may be of interest:

Assumption OS. $\chi(\bar{s}) \leq \bar{C} \left(1 + |\bar{s}|^{\bar{\lambda}}\right)$ for some positive constants \bar{C} and $\bar{\lambda}$.

Assumption SS. $\chi(\bar{s}) \leq \bar{C}_1 \left(1 + |\bar{s}|^{\bar{C}_2}\right) \exp\left(|\bar{s}|^{\bar{\lambda}}/\bar{C}_3\right)$ for some positive constants $\bar{C}_1, \bar{C}_2, \bar{C}_3$, and $\bar{\lambda}$, but $\chi(\bar{s})$ is not bounded by any polynomial in $|\bar{s}|$ on the real line.

Assumption OS and SS are generalizations of, correspondingly, the *ordinary-smooth* and *super-smooth* distribution assumptions made by Fan (1991). This classification is useful because the rates of convergence of the estimators of CDFs are polynomial in the sample size n when Assumption OS holds, but are only logarithmic in n when Assumption SS holds.

The following example is used throughout the section to illustrate the assumptions:

Example 1. Suppose that $U_{it} = \sigma_t(X_{it}) \xi_{it}$, where $\sigma_t(x)$ is the conditional heteroskedasticity function and ξ_{it} are i.i.d. (across i and t) random variables with probability distribution \mathcal{L}_ξ . Assume that $\inf_{x \in \mathcal{X}, 1 \leq t \leq T} \sigma_t(x) > 0$ and $\sup_{x \in \mathcal{X}, 1 \leq t \leq T} \sigma_t(x) < \infty$. In this case $\phi_{U_{it}}(s|x) = \phi_{\xi_{it}}(\sigma_t(x)s)$.

For instance, Assumption OS is satisfied for Example 1 when \mathcal{L}_ξ is a Laplace or Gamma distribution. In the former case, $\phi_{U_{it}}(s|x) = (1 + \sigma_t^2(x)s^2)^{-1}$, hence Assumption OS holds with $\bar{\lambda} = 2$. Similarly, when \mathcal{L}_ξ is a Gamma distribution with $\gamma > 0$ and $\theta > 0$ the conditional characteristic function has the form $\phi_{U_{it}}(s|x) = (1 - i\theta\sigma_t(x)s)^{-\gamma}$, and hence Assumption OS holds with $\bar{\lambda} = \gamma$.

When \mathcal{L}_ξ is normal in Example 1, Assumption SS is satisfied with $\bar{\lambda} = 2$ because $\phi_{U_{it}}(s|x) = \exp(-\sigma_t^2(x)s^2/2)$. Assumption SS is also satisfied with $\bar{\lambda} = 1$ when \mathcal{L}_ξ has Cauchy or Student-t distributions in Example 1.

Let $\alpha = (\alpha_1, \dots, \alpha_d)$ denote a d -vector of nonnegative integers and denote $|\alpha| = \alpha_1 + \dots + \alpha_d$. For any function $\varphi(\cdot)$ that depends on $z \in \mathcal{Z} \subset \mathbb{R}^d$ (and possibly some other variables)

define $\partial_z^\alpha \varphi(\cdot) = \partial^{|\alpha|} \varphi(\cdot) / \partial z_1^{\alpha_1} \dots \partial z_d^{\alpha_d}$. Function $\varphi : \mathcal{S} \times \mathcal{Z} \rightarrow \mathbb{C}$, $\mathcal{S} \subset \mathbb{R}$, $\mathcal{Z} \subset \mathbb{R}^d$ is said to belong to class of functions $\mathcal{D}_{\mathcal{Z}}^{\mathcal{S}}(k)$ if it is k times continuously differentiable in z and there is a constant $C > 0$ such that $\sup_{z \in \mathcal{Z}} \max_{|\alpha|=l} |\partial_z^\alpha \varphi(s, z)| \leq C(1 + |s|^l)$ for all $l \leq k$ and all $s \in \mathcal{S}$. The class of functions $\mathcal{D}_{\mathcal{Z}}^{\mathcal{S}}(k)$ is introduced because the k -th derivative of function $\phi_{m(x_t, \alpha_i)}(s|x_t)$ with respect to x_t contains terms multiplied by s^l , with l no bigger than k .

Assumption 4. For all t and some constant $\varsigma > 0$, $\phi_{U_{it}}(s|x_t) \in \mathcal{D}_{\mathcal{X}}^{\mathbb{R}}(\bar{k})$ and $\partial \phi_{U_{it}}(s|x_t) / \partial s \in \mathcal{D}_{\mathcal{X}}^{[-\varsigma, \varsigma]}(\bar{k})$.

Assumption 5. For all t and τ , $\phi_{m(x_t, \alpha_i)}(s|x_t) \in \mathcal{D}_{\mathcal{X}}^{\mathbb{R}}(\bar{k})$, $\phi_{m(x_t, \alpha_i)}(s|x_{t\tau}) \in \mathcal{D}_{\mathcal{X} \times \mathcal{X}}^{\mathbb{R}}(\bar{k})$, $\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau}) \in \mathcal{D}_{\mathcal{X} \times \mathcal{X}}^{\mathbb{R}}(\bar{k})$, $\partial \phi_{m(x_t, \alpha_i)}(s|x_t) / \partial s \in \mathcal{D}_{\mathcal{X}}^{[-\varsigma, \varsigma]}(\bar{k})$, and $\partial \phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau}) / \partial s \in \mathcal{D}_{\mathcal{X} \times \mathcal{X}}^{[-\varsigma, \varsigma]}(\bar{k})$.

The assumptions on the smoothness of functions $\phi_{m(x_t, \alpha_i)}(s|x_t)$, $\phi_{m(x_t, \alpha_i)}(s|x_{t\tau})$, and $\phi_{U_{it}}(s|x_t)$ are natural; these functions are conditional expectations that are estimated by usual kernel methods. The presence of the terms containing $s^{\bar{k}}$ in the \bar{k} -th order Taylor expansion of $\phi_{Y_{it}}(s|x_t)$ and $\phi_{U_{it}}(s|x_t)$ causes the bias of estimators $\hat{\phi}_{Y_{it}}(s|x_t)$ and $\hat{\phi}_{U_{it}}(s|x_t)$ to be proportional to $|s|^{\bar{k}}$ for large $|s|$.

To see why an assumption on smoothness of function $\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau})$ in $x_{t\tau}$ is needed, consider estimator $\hat{\phi}_U^S(s|x)$ based on (6). The value of $\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau})$ is close to unity when x_t is close to x_τ . Thus, one can obtain $\phi_{U_{it} - U_{i\tau}}(s|x_{t\tau})$ from the formula

$$\phi_{Y_{it} - Y_{i\tau}}(s|(X_{it}, X_{i\tau}) = x_{t\tau}) = \phi_{U_{it} - U_{i\tau}}(s|x_{t\tau}) \phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau}).$$

he assumption on smoothness of function $\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau})$ in $x_{t\tau}$ controls how fast $\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_{t\tau})$ approaches unity when the distance between x_t and x_τ shrinks, which is necessary to calculate the rate of convergence of $\hat{\phi}_U^S(s|x)$. Note also that in (13), $\hat{\phi}_{Y_{it}}(s|x_t)$ is divided by s while s passes through zero. Then, the analysis of $[\hat{\phi}_{Y_{it}}(s|x_t) - \phi_{Y_{it}}(s|x_t)]/s$ requires assumptions regarding the behavior of derivatives $\partial \phi_{m(x_t, \alpha_i)}(s|x_t) / \partial s$ and $\partial \phi_{U_{it}}(s|x_t) / \partial s$ in the neighborhood of $s = 0$ because $\phi_{Y_{it}}(s|x_t) = \phi_{m(x_t, \alpha_i)}(s|x_t) \phi_{U_{it}}(s|x_t)$.

It is straightforward to check that Assumption 4 is satisfied in Example 1 if $\sigma_t(x) \in \mathcal{D}_{\mathcal{X}}(\bar{k})$ for all t and \mathcal{L}_{ξ} is Normal, Gamma, Laplace, or Student-t (with two or more degrees of freedom).

Sufficient conditions to ensure that Assumptions 4 and 5 hold are provided by Lemmas 2 and 3, respectively.

Lemma 2. Suppose there is a positive integer k , a constant C and a function $M(u, x)$ such that $f_{U_{it}}(u|x_t)$ has k continuous mixed partial derivatives in x_t and $\max_{|\alpha| \leq k} |\partial_{x_t}^\alpha f_{U_{it}}(u|x_t)|^2 / f_{U_{it}}(u|x_t) \leq M(u, x_t)$ for all $x_t \in \mathcal{X}$ and for almost all $u \in \mathbb{R}$, and $\int_{-\infty}^{\infty} M(u, x) du \leq C$ for all $x \in \mathcal{X}$. Suppose also that the support of U_{it} , conditional on $X_{it} = x_t$, is the whole real line

(and does not depend on x_t) and that $E[U_{it}^2 | X_{it} = x_t] \leq C$. Then Assumption 4 is satisfied with $\bar{k} = k$.

For instance, the conditions of the lemma are easily satisfied by Example 1 when $\sigma_t(x)$ has \bar{k} bounded derivatives for all t and all $x \in \mathcal{X}$, and \mathcal{L}_ξ is Normal, Gamma, Laplace, or Student-t (with three or more degrees of freedom).

The idea behind the conditions of the lemma is the following. The conditional characteristic function is defined as $\phi_{U_{it}}(s|x_t) = \int_{\text{Supp}(U_{it})} \exp\{isu\} f_{U_{it}|X_{it}}(u|x_t) du$, where $\text{Supp}(U_{it})$ is the support of U_{it} . Note that $|\exp\{isu\}| = 1$. Suppose that for some $k \geq 1$, all mixed partial derivatives of function $f_{U_{it}|X_{it}}(u|x_t)$ in x_t up to the order k are bounded. Also, suppose for a moment that $\text{Supp}(U_{it})$ is bounded. Then all mixed partial derivatives of function $\phi_{U_{it}}(s|x_t)$ in x_t up to the k -th order are bounded. This is because for any α such that $|\alpha| \leq k$ one has $|\partial_{x_t}^\alpha \phi_{U_{it}}(s|x_t)| \leq \left(\int_{\text{Supp}(U_{it})} du \right) |\partial_{x_t}^\alpha f_{U_{it}|X_{it}}(u|x_t)|$. However, this argument fails when the support of U_{it} is unbounded since $\int_{\text{Supp}(U_{it})} du = \infty$. Thus, when U_{it} has unbounded support one needs to impose some assumption about the relative sensitivity of the density $f_{U_{it}|X_{it}}(u|x_t)$ to changes in x_t compared to the magnitude of $f_{U_{it}|X_{it}}(u|x_t)$ in the tails. This is the main assumption of Lemma 2.

Similar sufficient conditions can be given for Assumption 5 to hold:

Lemma 3. Denote $\omega(s, x_t, \alpha) = m(x_t, \alpha) e^{ism(x_t, \alpha)}$. Suppose that functions $m(x_t, \alpha)$, $f_{\alpha_i}(\alpha|x_t)$, and $f_{\alpha_i}(\alpha|x_{t\tau})$ have k continuous mixed partial derivatives in, correspondingly, x_t , x_t , and $x_{t\tau}$ for all α and $x_{t\tau} \in \mathcal{X} \times \mathcal{X}$. Suppose that $\max_{|\alpha| \leq k} E[|\partial_{x_t}^\alpha \omega(s, x_t, \alpha_i)|^2 | x_{t\tau}] \leq C(1 + s^k)^2$ for some constant C and for all $s \in \mathbb{R}$ and $x_{t\tau} \in \mathcal{X} \times \mathcal{X}$. Suppose the support of $f_{\alpha_i}(\alpha|x_t)$ is $(\psi_l(x_t), \psi_h(x_t))$, and that for each $j \in \{l, h\}$, t , and $x_t \in \mathcal{X}$ either (a) $\psi_j(x_t)$ is infinite, or (b) $\psi_j(x_t)$ has k continuous mixed partial derivatives in x_t and $\max_{|\alpha| \leq k-1} |\partial_{x_t}^\alpha f_{\alpha_i}(\alpha|x_t)| \Big|_{\alpha = \psi_j(x_t)} = 0$. Suppose that $\max_{|\alpha| \leq k} (\partial_{x_t}^\alpha f_{\alpha_i}(\alpha|x_t))^2 / f_{\alpha_i}(\alpha|x_t) \leq M(\alpha, x_t)$ for all $\alpha \in (\psi_l(x_t), \psi_h(x_t))$, where $\int_{\psi_l(x_t)}^{\psi_h(x_t)} M(\alpha, x_t) d\alpha < C$ for all $x_t \in \mathcal{X}$. Finally, assume that analogous restrictions hold for $f_{\alpha_i}(\alpha|x_{t\tau})$. Then Assumption 5 is satisfied with $\bar{k} = k$.

Lemma 3 extends Lemma 2 by allowing the conditional (on x_t) support of random variable (α_i) to depend on x_t . This extension requires assumptions on smoothness of support boundary functions $\psi_l(x_t)$ and $\psi_h(x_t)$ in x_t .

Assumption 6. Multivariate kernel $K(v)$ has the form $K(v) = \prod_{l=1}^p \tilde{K}(v_l)$, where $\tilde{K}(\xi)$ is a univariate \bar{k} -th order kernel function that satisfies $\int |\xi|^{\bar{k}} \tilde{K}(\xi) d\xi < \infty$, $\int \tilde{K}^2(\xi) d\xi < \infty$, and $\sup_{\xi \in \mathbb{R}} |\partial \tilde{K}(\xi) / \partial \xi| < \infty$.

Assumption 7. (i) $K_w(\cdot)$ is a symmetric (univariate) kernel function whose Fourier transform $w(\cdot)$ has support $[-1, 1]$. (ii) $w(s) = 1 + O(|s|^{k_F})$ as $s \rightarrow 0$.

Assumption 7(ii) implies that $K_w(\cdot)$ is a k_F -th order kernel. For instance, one can take $w(s) = (1 - |s|^{k_F})^r I(|s| \leq 1)$ for some integer $r \geq 1$.²⁶

Assumption SYM. $\phi_{U_{it}}(s|x) = \phi_{U_{it}}(-s|x) = \phi_{U_{i\tau}}(s|x)$ for all t, τ , and all $(s, x) \in \mathbb{R} \times \mathcal{X}$.

Assumption SYM is satisfied when U_{it} have the same symmetric conditional distribution for all t . The estimator $\hat{\phi}_{U_{it}}^S(s|x)$ is consistent when Assumption SYM holds and can be plugged into the estimator $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$. If one is unwilling to impose Assumption SYM the following assumption is useful:

Assumption ASYM. (i) There exist constants $C_\phi > 0$ and $b_\phi \geq 0$ such that for all $s \in \mathbb{R}$ it holds that

$$\sup_{x \in \mathcal{X}} |\partial \ln \phi_{U_{it}}(s|x) / \partial s| \leq C_\phi \left(1 + s^{b_\phi}\right),$$

and $b_\phi = 0$ when OS holds; (ii) function $E[m(x_t, \alpha_i) e^{is(m(x_t, \alpha_i) - m(x_\tau, \alpha_i))} | x_{t\tau}]$ belongs to $\mathcal{D}_{\mathcal{X} \times \mathcal{X}}^{\mathbb{R}}(\bar{k})$ and $\partial \phi_{U_{it}}(s|x_t) / \partial s \in \mathcal{D}_{\mathcal{X}}^{\mathbb{R}}(\bar{k})$.

Assumption ASYM is very mild. For instance, in Example 1 ASYM(i) is satisfied for \mathcal{L}_ξ being Gamma, Laplace, Normal, or Student-t distributions. Imposing the condition $b_\phi = 0$ appears not to be restrictive for ordinary smooth distributions. The rates of convergence can be derived without ASYM(i) using the results of Theorem 3 below; however, this case seems to be of little practical interest and its analysis is omitted. ASYM(ii) is a very minor extension of Assumptions 4 and 5. The conditions of Lemma 3 are sufficient for ASYM(ii) to hold. Below, for notational convenience, take $b_\phi = 0$ if Assumption SYM holds.

In the results below it is implicitly assumed that the estimator $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$ uses $\hat{\phi}_{U_{it}}^S(s|x_t)$ as an estimator of $\phi_{U_{it}}(s|x_t)$ when Assumption SYM holds. When Assumption ASYM holds, $\hat{\phi}_{U_{it}}^{AS}(s|x_t)$ is used to estimate $\phi_{U_{it}}(s|x_t)$ in the estimator $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$. Thus, the derived rates of convergence below are obtained for two different estimators of $\phi_{U_{it}}(s|x_t)$. The estimator $\hat{\phi}_{U_{it}}^S(s|x_t)$ requires symmetry, but its rate of convergence is (slightly) faster than the rate of convergence of estimator $\hat{\phi}_{U_{it}}^{AS}(s|x_t)$.

Note that the above assumptions allow the characteristic functions to be complex-valued (under Assumption ASYM) and do not require that the characteristic functions have bounded support.

Suppose that the bandwidths $h_Y(p)$, $h_Y(2p)$, h_U , and h_w satisfy the following assumption:²⁷

²⁶Delaigle and Hall (2006) consider the optimal choice of kernel in the unconditional density deconvolution problem. They find that the kernel with Fourier transform $w(s) = (1 - s^2)^3 I(|s| \leq 1)$ performs well in their setting.

²⁷In principle one can allow the bandwidths $h_Y(d)$ and h_U to depend on s . This may improve the rate of convergence of the considered estimators in some special cases, although it does not improve the rates

Assumption 8. For $d \in \{p, 2p\}$ the following hold: (i) $\min\{nh_U, nh_Y(d)\} \rightarrow \infty$, $\max\{h_U, h_Y(d), h_w\} \rightarrow 0$, (ii) $[\log(n)]^{1/2} n^{1/\gamma-1/2} \max\{h_U^{-p}, [h_Y(d)]^{-d/2}\} \rightarrow 0$, (iii) $(\log(n))^{\gamma-3/2} h_w^{-1} \max\{h_U^p, [h_Y(d)]^{d/2}\} n^{1/\gamma-1/2} \rightarrow 0$, (iv) $([\log(n)/(nh_U^{2p})]^{1/2} + h_w^{-\bar{r}} h_U^{\bar{k}}) h_w^{-b_\phi-1} \times \chi^2(h_w^{-1}) \rightarrow 0$, and (v) if ASYM holds $h_Y(p) = O(h_U/h_w)$ and $h_U = o(h_Y^{1/2})$.

Assumptions 8(ii) and (iii) are used in conjunction with Bernstein's inequality and essentially require that the tail of Y_{it} is sufficiently thin compared to the rate of convergence below. Notice that the larger γ in Assumption 3 is, the easier it is to satisfy Assumption 8. Assumption 8(iv) ensures that $[\hat{\phi}_{U_{it}}(s|x_t) - \phi_{U_{it}}(s|x_t)] / \phi_{U_{it}}(s|x_t) = o_p(1)$, which is necessary because of the term $\hat{\phi}_{U_{it}}(s|x_t)$ in the denominator in the formula for $\hat{F}_{m(x_t, \alpha_i)}(\cdot)$. Assumption 8(v) is very mild and assures that the second fraction in the definition of estimator $\hat{\phi}_{U_{it}}^{AS}(\cdot)$ does not dominate the first one. Assumptions 8(i) and (iv) are necessary for consistency. On the other hand, Assumptions 8(ii),(iii), and (v) are not necessary for obtaining consistency and the rate of convergence can be derived without imposing these assumptions. However, the resulting expression for the rate of convergence is more complicated.

The following theorem gives $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ and $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$.

Theorem 3. Suppose Assumptions ID(i-vi), 1-8, and either SYM or ASYM hold. Then,

$$\begin{aligned} \sup_{(\omega, x_{t\tau}) \in \mathbb{R} \times \bar{\mathcal{X}} \times \bar{\mathcal{X}}} \left| \hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau}) - F_{m(x_t, \alpha_i)}(\omega|x_{t\tau}) \right| &= O_p(\beta_n(2p)) \text{ and} \\ \sup_{(\omega, x_t) \in \mathbb{R} \times \bar{\mathcal{X}}} \left| \hat{F}_{m(x_t, \alpha_i)}(\omega|x_t) - F_{m(x_t, \alpha_i)}(\omega|x_t) \right| &= O_p(\beta_n(p)), \end{aligned}$$

where

$$\begin{aligned} \beta_n(d) &= h_w^{k_F} + \left(\left[\log(n) / (nh_Y^d(d)) \right]^{1/2} + h_w^{-(\bar{k}-1)} h_Y^{\bar{k}}(d) \right) h_w^{-1} \chi(h_w^{-1}) \\ &\quad + \left([\log(n)/(nh_U^{2p})]^{1/2} + h_w^{-\bar{r}} h_U^{\bar{k}} \right) \max \left\{ 1, \bar{h}_w^{k_F-1-b_\phi} \chi^2(h_w^{-1}) \right\}, \end{aligned}$$

where $\bar{r} = \bar{k} - 1$ under SYM and $\bar{r} = \bar{k}$ under ASYM, and the set $\bar{\mathcal{X}}$ is any set that satisfies $\bar{\mathcal{X}} + B_{\varepsilon_{\mathcal{X}}}^p(0) \subset \mathcal{X}$, where $B_{\varepsilon_{\mathcal{X}}}^p(0)$ is a p -dimensional ball around 0^p with an arbitrary small radius $\varepsilon_{\mathcal{X}} > 0$.²⁸

The rate of convergence $\beta_n(d)$ has three distinctive terms. The first term is the regularization bias term. This bias is present because (13) is the inverse Fourier transform of $\phi_w(s) \hat{\phi}_m(s|\cdot)$, the regularized version of $\hat{\phi}_m(s|\cdot)$, and not of the true characteristic function of convergence in the leading cases. Allowing the bandwidths to depend on s is also unappealing from the practical standpoint, since one would need to provide "bandwidth functions" $h_U(s)$ and $h_Y(d, s)$, which appears impractical.

²⁸Here set $\bar{\mathcal{X}} + B_{\varepsilon_{\mathcal{X}}}^p(0)$ is the set of points $\{x = x_1 + x_2 : x_1 \in \bar{\mathcal{X}}, x_2 \in B_{\varepsilon_{\mathcal{X}}}^p(0)\}$

$\phi_m(s|\cdot)$. The second and the third terms of $\beta_n(d)$ contain parentheses, which are the familiar expressions for the rate of convergence of kernel estimators of $\hat{\phi}_Y(s|\cdot)$ and $\hat{\phi}_{U_{it}}(s|x_t)$, correspondingly. The h_w^{-1} and $\chi(h_w^{-1})$ parts of the second term of $\beta_n(d)$ come from the integration over s and division by $\hat{\phi}_{U_{it}}(s|x_t)$, respectively. The third term of $\beta_n(d)$ is the error from the estimation of $\hat{\phi}_{U_{it}}(s|x_t)$, i.e. essentially the error from estimation of the integral operator in the ill-posed inverse problem. While the term $\hat{\phi}_{U_{it}}(s|x_t)$ magnifies estimation error for large s , larger k_F helps reducing this error. Yet, these effects play a role only for large s , and not for small s (the rate of convergence of $\hat{\phi}_{U_{it}}(s|x_t)$ for s of order one does not depend on k_F or the function $\chi(\bar{s})$). This is the logic behind the $\max\{\dots\}$ part of the third term of $\beta_n(d)$; see also Remark 11 below.

Parameters d and \bar{r} are introduced because the result of Theorem 3 covers four different estimators: estimators $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ and $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ (corresponding to $d = 2p$ and $d = p$, respectively), each using $\hat{\phi}_{U_{it}}(s|x) = \hat{\phi}_{U_{it}}^S(s|x)$ (when Assumption SYM holds; $\bar{r} = \bar{k}$) or $\hat{\phi}_{U_{it}}(s|x) = \hat{\phi}_{U_{it}}^{AS}(s|x)$ (when Assumption ASYM holds; $\bar{r} = \bar{k} - 1$). Assumption ASYM is more general than Assumption SYM. However, the corresponding estimator $\hat{\phi}_{U_{it}}^{AS}(s|x)$ contains integration, while estimator $\hat{\phi}_{U_{it}}^S(s|x)$ does not. As a result, the rate of convergence of estimator $\hat{\phi}_{U_{it}}^S(s|x)$ is slightly faster than the rate of convergence of estimator $\hat{\phi}_{U_{it}}^{AS}(s|x)$. This difference in the rates of convergence is reflected in the rate of convergence $\beta_n(d)$.

The sole purpose of introducing set $\bar{\mathcal{X}}$ in the statement of the theorem is to avoid considering estimation at the boundary points of the support of X_{it} , where the estimators are inconsistent. For example, when $\mathcal{X} = [a, b]$ one can take $\bar{\mathcal{X}} = [a + \varepsilon_{\mathcal{X}}, b - \varepsilon_{\mathcal{X}}]$, $\varepsilon_{\mathcal{X}} > 0$. In fact, it is easy to modify the estimators $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ and $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ so that the rate of convergence result will hold for all $x_{t\tau} \in \mathcal{X} \times \mathcal{X}$. For instance, one can follow the boundary kernel approach, see, for example, Müller (1991).

The next theorem shows what the rate of convergence $\beta_n(d)$ becomes when the disturbances U_{it} have an ordinary smooth distribution and the bandwidths h_w , h_U , and $h_Y(d)$ are chosen optimally. Define

$$\begin{aligned} \tilde{k}_F(d) &= 1 + \bar{\lambda} + [2\bar{\lambda}(2p - d) + 2p\bar{r} - d\bar{k} + 4p - d] / (2\bar{k} + d) \quad \text{and} \\ \check{k}_F &= \bar{\lambda}(2\bar{k}/p + 1) + \bar{k} - 2 - 2\bar{r}, \end{aligned}$$

and note that for $d = 2p$ under Assumption SYM, $\tilde{k}_F(d)$ simplifies to $\tilde{k}_F(2p) = 1 + \bar{\lambda}$. The following assumption is a sufficient condition to ensure that Assumption 8 is satisfied by the optimal bandwidths h_w , h_U , and $h_Y(d)$.

Assumption 9. (i) $\gamma > 6(p + 1)(p + 2)/(p + 6)$; (ii) if $p \geq 3$ then, in addition, $k_F \geq \bar{k}$ when $k_F \geq \tilde{k}_F(d)$, and $\bar{\lambda} \geq \bar{k}$ when $k_F < \tilde{k}_F(d)$.²⁹

²⁹The above assumption is only a sufficient condition. The restriction on the number of bounded moments

Theorem 4. *Suppose Assumption OS holds with $\bar{\lambda} \geq 1$, and also Assumptions ID(i-vi), 1-7, 9, and either SYM or ASYM hold. Then, with the appropriately chosen bandwidths,*

$$\begin{aligned} \sup_{(\omega, x_{t\tau}) \in \mathbb{R} \times \bar{\mathcal{X}} \times \bar{\mathcal{X}}} \left| \hat{F}_{m(x_t, \alpha_i)}(\omega | x_{t\tau}) - F_{m(x_t, \alpha_i)}(\omega | x_{t\tau}) \right| &= O_p(\beta_n^{OS}(2p)) \text{ and} \\ \sup_{(\omega, x_t) \in \mathbb{R} \times \bar{\mathcal{X}}} \left| \hat{F}_{m(x_t, \alpha_i)}(\omega | x_t) - F_{m(x_t, \alpha_i)}(\omega | x_t) \right| &= O_p(\beta_n^{OS}(p)), \end{aligned}$$

where

$$\begin{aligned} \beta_n^{OS}(2p) &= \begin{cases} (\log(n)/n)^{\rho_n^U}, & \text{when } k_F < \tilde{k}_F(2p) \text{ and } k_F < 2\bar{\lambda} + 1, \\ (\log(n)/n)^{\rho_n(2p)}, & \text{otherwise,} \end{cases} \\ \beta_n^{OS}(p) &= \begin{cases} (\log(n)/n)^{\rho_n^U}, & \text{when } k_F \leq \tilde{k}_F(p) \text{ and } k_F \leq 2\bar{\lambda} + 1, \\ (\log(n)/n)^{\rho_n(p)}, & \text{when } k_F > \tilde{k}_F(p) \text{ and } k_F \leq 2\bar{\lambda} + 1, \\ (\log(n)/n)^{\rho_n(p)}, & \text{when } k_F \leq \check{k}_F \text{ and } k_F > 2\bar{\lambda} + 1, \\ (\log(n)/n)^{\rho_n^U}, & \text{when } k_F > \check{k}_F \text{ and } k_F > 2\bar{\lambda} + 1, \end{cases} \\ \rho_n(d) &= \frac{1}{2} \frac{k_F}{k_F + 1 + \bar{\lambda} + (\bar{k} - 1)d / (2\bar{k} + d)} \frac{2\bar{k}}{2\bar{k} + d}, \\ \rho_n^U &= \frac{1}{2} \frac{k_F}{\max\{k_F, 2\bar{\lambda} + 1\} + 1 + \bar{r}p / (\bar{k} + p)} \frac{2\bar{k}}{2\bar{k} + 2p}. \end{aligned} \tag{15}$$

The corresponding values of h_U , $h_Y(d)$, and h_w are given in the proof of the theorem.

Remark 10. *The rate of convergence (15) is intuitive. If one (formally) takes $d = 0$, the rate of convergence becomes $k_F / (2k_F + 2 + 2\bar{\lambda})$, which is the rate of convergence (except for the $\log(n)$ factor) obtained by Fan (1991) for the unconditional deconvolution problem with known distribution of errors U_{it} (see his Remark 3). On the other hand, suppose function $F_{m(x_t, \alpha_i)}(\omega | x_t)$ is very smooth with respect to ω , i.e. suppose that $k_F \rightarrow \infty$ while all the other parameters in (15) are fixed. In this case $\rho_n(d) \rightarrow \bar{k} / (2\bar{k} + d)$, which is the rate of convergence obtained by Stone (1982) for nonparametric regression/density estimation. Thus, formula (15) combines conditional (Stone, 1982) and deconvolution (Fan, 1991) aspects of considered problem.*

Remark 11. *Some discussion of the different cases for the rate of convergence in Theorem 4 is in order. The error in estimation of $F_{m(x_t, \alpha_i)}(\cdot)$ comes from the estimation error of the estimators $\hat{\phi}_{Y_{it}}(\cdot)$ and $\hat{\phi}_{U_{it}}(\cdot)$. It will be convenient to write $\beta_n(d)$ as $\beta_n(d) = h_w^{k_F} + T_Y(d) +$*

γ in fact corresponds to the least smooth case $k_F = 1$; the necessary number of bounded moments γ can be reduced if $F_m(w|\cdot)$ has more than one bounded derivative in w (i.e. $k_F > 1$). Condition (ii) is technical and does not appear to be restrictive. Note that condition $k_F < \tilde{k}_F(d)$ means that $\bar{\lambda}$ is relatively larger than k_F , therefore it appears that the condition $\bar{\lambda} \geq \bar{k}$ imposed in this case is mild.

T_U , where the terms $T_Y(d)$ and T_U correspond to the stochastic errors of estimation of $\phi_{Y_{it}}(\cdot)$ and $\phi_{U_{it}}(\cdot)$, respectively. The rate of convergence $\rho_n(d)$ corresponds to the case when $T_Y(d)$ is larger than T_U (asymptotically). That is, the rate of convergence $\rho_n(d)$ (and the implied optimal bandwidth h_w) balances h_w^F and $T_Y(d)$. Since the estimation error of $\hat{\phi}_{U_{it}}(s|x_t)$ is smaller in this case, the rate of convergence $\rho_n(d)$ is the same as in the hypothetical case when the conditional distribution of U_{it} were known and $\hat{\phi}_{U_{it}}(s|x_t)$ replaced by $\phi_{U_{it}}(s|x_t)$ in (13). On the other hand, when T_U dominates $T_Y(d)$, the rate of convergence is ρ_n^U (i.e. ρ_n^U balances $h_w^{k_F}$ and T_U).

Consider the estimator $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$. When k_F is small, the estimation error is given by the error of estimation of $\phi_{U_{it}}(s|x_t)$ (and the rate of convergence of $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ is ρ_n^U). Note that larger k_F reduces T_U , but does not affect $T_Y(2p)$. Thus, for sufficiently large k_F the term T_U becomes small and the rate of convergence of the estimator $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_{t\tau})$ becomes $\rho_n(2p)$.

The situation is slightly more complicated for the estimator $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$. As in the previous case, the estimation error is dominated by T_U when k_F is small, but is dominated by $T_Y(p)$ for larger k_F . However, in contrast to the previous case, for very large k_F the rate of convergence is again determined by T_U . To see why this is the case, note that for $d = p$ the term $T_Y(p)$ is "p-dimensional", i.e. the rate of convergence in the parentheses of the second term $\beta_n(p)$ corresponds to the estimation of a conditional mean, given $X_{it} = x_t$, where X_{it} is p-dimensional. In contrast, the parentheses of the third term of $\beta_n(p)$ correspond to the estimation of a conditional mean, given $X_{it} = X_{i\tau} = x_t$, thus the third term of $\beta_n(p)$ is "2p-dimensional" and the parentheses are converging to zero at a slower rate than the parentheses in the second term of $\beta_n(p)$. Thus, higher k_F helps reducing the term T_U , but only to some limit (because of the $\max\{\dots\}$ term in T_U); thus for $k_F > \check{k}_F$ the T_U dominates $T_Y(p)$ and the rate of convergence of the estimator $\hat{F}_{m(x_t, \alpha_i)}(\omega|x_t)$ is ρ_n^U .³⁰

Remark 12. Note that the results developed in the paper may also be applied to the pure conditional deconvolution model $Y_{it} = \alpha_i + U_{it}$ (or more generally $Y_{it} = m(X_{it}) + \alpha_i + U_{it}$), where α_i and U_{it} are independent conditional on X_{it} . In this model the estimators do not change, although the rate of convergence improves. The rate of convergence is $\beta_n^{OS}(d)$, but the terms $(\bar{k} - 1)d / (2\bar{k} + d)$ and $\bar{r}p / (2\bar{k} + p)$ in the denominator of the first fractions (the "deconvolution" fractions) are absent. Model (1) subjects α_i to transformation $m(x, \cdot)$, which is responsible for the terms $(\bar{k} - 1)d / (2\bar{k} + d)$ and $\bar{r}p / (2\bar{k} + p)$ appearing in the rate $\beta_n^{OS}(d)$ of Theorem 4 as well as the terms $h_w^{-(\bar{k}-1)}$ and $h_w^{-\bar{r}}$ appearing in rate $\beta_n(d)$ of Theorem 3. These terms do not appear in the corresponding rates of convergence in model $Y_{it} = \alpha_i + U_{it}$.

Now consider the case of (conditionally) super-smooth errors.

³⁰ As usual, the larger k_F is, the slower is the rate of convergence of h_w to zero, because it is easier to control the regularization bias $h_w^{k_F}$ for larger k_F .

Assumption 10. $\gamma > 2 + 2p$.

Theorem 5. *Suppose Assumptions SS, ID(i-vi), 1-7, 10, and either SYM or ASYM hold. Then, with the appropriately chosen bandwidths,*

$$\begin{aligned} \sup_{(\omega, x_{t\tau}) \in \mathbb{R} \times \bar{\mathcal{X}} \times \bar{\mathcal{X}}} \left| \hat{F}_{m(x_t, \alpha_i)}(\omega | x_{t\tau}) - F_{m(x_t, \alpha_i)}(\omega | x_{t\tau}) \right| &= O_p(\beta_n^{SS}) \text{ and} \\ \sup_{(\omega, x_t) \in \mathbb{R} \times \bar{\mathcal{X}}} \left| \hat{F}_{m(x_t, \alpha_i)}(\omega | x_t) - F_{m(x_t, \alpha_i)}(\omega | x_t) \right| &= O_p(\beta_n^{SS}), \end{aligned}$$

where $\beta_n^{SS} = (\ln(n))^{-k_F/\bar{\lambda}}$. The corresponding values of h_U , $h_Y(d)$, and h_w are given in the proof of the theorem.

Remark 13. *The rate of convergence is logarithmic and does not depend on the dimension of the vector of the covariates X_{it} . This is similar to the nonparametric instrumental variable estimation in the severely ill-posed case, see, for example, Chen and Reiss (2007).*

Remark 14. *One may also be interested in the pointwise (in x_t) rate of convergence of $\hat{F}_{m(x_t, \alpha_i)}(s | x_t)$. In this case the definition of $\chi(\bar{s})$ should be changed to $\chi(\bar{s} | x) = \max_{1 \leq t \leq T} \sup_{(s) \in [-\bar{s}, \bar{s}]} 1/|\phi_{U_{it}}(s | x)|$, and the Assumptions OS and SS should be changed accordingly, so that $\bar{\lambda}$ becomes a function of x_t . Then the pointwise rates of convergence are the same as the rates given in Theorems 3, 4, and 5 with $\bar{\lambda}$ substituted by $\bar{\lambda}(x_t)$, as long as the function $\bar{\lambda}(x)$ is sufficiently smooth in x .*

Corollary 6. *Suppose there is a set $\Omega_{x\omega} = \{(x_t, \omega) \in \bar{\mathcal{X}} \times \mathbb{R} : \omega_a(x_t) \leq \omega \leq \omega_b(x_t)\}$ and positive numbers $\varepsilon \leq \varepsilon_{\mathcal{X}}$ and δ such that $f_{m(\alpha_i, x_t)}(\omega | x_t) \geq \delta$ for all $(x_t, \omega) \in \bar{\Omega}_{x\omega} + B_{\varepsilon_\omega}^{p+1}(0)$, where $B_{\varepsilon_\omega}^{p+1}(0)$ is a $p + 1$ -dimensional ball around 0^{p+1} with radius ε_ω . Suppose that the conditions of either Theorem 3, or 4, or 5 hold. Then*

$$\sup_{(q, x_t) \in \left[F_{m(x_t, \alpha_i)}(\omega_a(x_t) | x_t), F_{m(x_t, \alpha_i)}(\omega_b(x_t) | x_t) \right] \times \bar{\mathcal{X}}} \left| \hat{Q}_{m(x_t, \alpha_i)}(q | x_t) - Q_{m(x_t, \alpha_i)}(q | x_t) \right| = O_p(\psi_n),$$

where ψ_n is either $\beta_n(p)$, or $\beta_n^{OS}(p)$, or $\beta_n^{SS}(p)$, respectively.

A similar corollary holds for the estimator $\hat{Q}_{m(x_t, \alpha_i)}(q | x_{t\tau})$. The derived rates of convergence can be used to obtain the rates of convergence of estimators \hat{m}_{RE} and \hat{m}_{FE} .

Theorem 7. *Suppose Assumption RE holds. Suppose that for some small $\delta > 0$ function $\partial m(x, \alpha) / \partial \alpha$ is bounded away from zero and infinity for all $\alpha \in [\delta/2, 1 - \delta/2]$ and all $x \in \mathcal{X}$. Suppose that the conditions of either Theorem 3, or 4, or 5 hold. Then*

$$\sup_{(x, \alpha) \in \bar{\mathcal{X}} \times [\delta, 1 - \delta]} |\hat{m}_{RE}(x, \alpha) - m(x, \alpha)| = O_p(\beta_n^{RE}),$$

where β_n^{RE} is either $\beta_n(p)$, or $\beta_n^{OS}(p)$, or β_n^{SS} , respectively.

Now consider the fixed effects model. Define functions

$$\begin{aligned}\underline{\alpha}(q|x) &= \min_{\{(t,\tau):t\neq\tau\}} \inf_{x_2\in\mathcal{X}} Q_{\alpha_i|X_{it},X_{i\tau}}(q|x,x_2) \text{ and} \\ \bar{\alpha}(q|x) &= \max_{\{(t,\tau):t\neq\tau\}} \sup_{x_2\in\mathcal{X}} Q_{\alpha_i|X_{it},X_{i\tau}}(q|x,x_2).\end{aligned}$$

Also, define the set $S_{x,\alpha}(\vartheta) = \{(x,\alpha) : x \in \bar{\mathcal{X}}, \alpha \in [\underline{\alpha}(\vartheta|x), \bar{\alpha}(1-\vartheta|x)]\}$.

Theorem 8. *Suppose Assumption FE holds. Suppose that for some small $\delta > 0$ functions $f_{\alpha_i|X_{it}}(\alpha|x_t)$, $f_{\alpha_i|X_{it},X_{i\tau}}(\alpha|x_t,x_\tau)$ and $\partial m(x_t,\alpha)/\partial\alpha$ are bounded away from zero and infinity for all $\alpha \in [\underline{\alpha}(\delta/2|x_t), \bar{\alpha}(1-\delta/2|x_t)]$, $(x_t,x_\tau) \in \mathcal{X} \times \mathcal{X}$, t , and $\tau \neq t$. Suppose that the conditions of either Theorem 3, or 4, or 5 hold. Then for all t ,*

$$\begin{aligned}\sup_{(x,\alpha)\in S_{x,\alpha}(\delta)} |\hat{m}_{FE}(x,\alpha) - m(x,\alpha)| &= O_P(\beta_n^{FE}), \\ \sup_{(x,\alpha)\in\bar{\mathcal{X}}\times[\delta,1-\delta]} \left| \hat{Q}_{\alpha_i|X_{it}}(q|x_t) - Q_{\alpha_i|X_{it}}(q|x_t) \right| &= O_P(\beta_n^{FE}),\end{aligned}$$

where β_n^{FE} is either $\beta_n(2p)$, or $\beta_n^{OS}(2p)$, or β_n^{SS} , respectively.

Suppose the conditions of Theorem 3 hold. Then, the rate of convergence of $\hat{m}_{RE}(\cdot)$ to $m(\cdot)$ is $\beta_n(p)$, i.e. the rate of convergence of estimator $\hat{m}_{RE}(\cdot)$ to $m(\cdot)$ is " p -dimensional". This is because estimation of the random effects model only requires estimation of cumulative distribution and quantile functions conditional on the p -vector x_t , i.e. it uses between-variation. In contrast, the estimation of the fixed effects model relies on within-variation and hence requires estimation of distribution and quantile functions conditional on the $(2p)$ -vector $x_{t\tau}$. Consequently, Theorem 3 ensures that the rate of convergence of the estimator $\hat{m}_{FE}(\cdot)$ to $m(\cdot)$ is at least $\beta_n(2p)$. Importantly, estimator $\hat{m}_{FE}(\cdot)$ has the form of averaging over x_2 (see equation (14)). Thus, it is likely that the rate of convergence of estimator $\hat{m}_{FE}(\cdot)$ to $m(\cdot)$ is faster than $\beta_n(2p)$ under some conditions. One way of showing this improvement in the rate of convergence would be to follow Newey (1994) in deriving the influence of each observation. However, the estimation problem in this paper is ill-posed and the analysis would require a substantial deviation from Newey (1994). This analysis is left for future research.

Finally, the estimator of the policy relevant function $\hat{h}(x,q)$ has the same rate of convergence as the estimator $\hat{m}_{FE}(x,\alpha)$. Note that estimation of $h(x,q)$ requires estimation of the unconditional distribution of α_i . For instance, estimators $\hat{F}_{\alpha_i}(\alpha) = \int_{\mathcal{X}} \hat{F}_{\alpha_i}(\alpha|x_t) \hat{f}(x_t) dx_t$ or $\hat{F}_{\alpha_i}(\alpha) = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{F}_{\alpha_i}(\alpha|X_{it})$ can be used. Note that these estimators use the values of $\hat{F}_{\alpha_i}(\alpha|x_t)$ for all values $x_t \in \mathcal{X}$, thus one may need to modify the estimator $\hat{F}_{\alpha_i}(\alpha|x_t)$ so that it is consistent on the boundary of set \mathcal{X} . For example, one can use boundary kernels, as was mentioned earlier.

4 Monte Carlo

This section presents a Monte Carlo study of the finite sample properties of the estimator $\hat{h}(x, q)$ for the fixed effects model. The data is generated according to the following model:

$$\begin{aligned} Y_{it} &= m(X_{it}, \alpha_i) + U_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T = 2, \\ m(x, \alpha) &= 2\alpha + (2 + \alpha)(2x - 1)^3, \\ X_{it} &\sim i.i.d. \text{ Uniform } [0, 1], \\ \alpha_i &= \frac{\rho}{\sqrt{T}} \sum_{t=1}^T \sqrt{12} (X_{it} - 0.5) + \sqrt{1 - \rho^2} \psi_i, \quad \rho = 0.5, \\ U_{it} &= (1 + X_{it}) \sigma_0 \varepsilon_{it}, \end{aligned}$$

where $n \in \{2500, 10000\}$ and the following distributions of ψ_i are considered:

- Design I: $\psi_i \sim i.i.d. N(0, 1)$;
- Design II: $\psi_i \sim i.i.d. (\chi^2(4) - 4) / \sqrt{8}$;³¹
- Design III:

$$\psi_i \sim i.i.d. \begin{cases} N(-2.5, 1) / \sqrt{29/4}, & \text{with probability } 1/2, \\ N(2.5, 1) / \sqrt{29/4} & \text{with probability } 1/2. \end{cases}$$

The distribution of ψ_i is symmetric, skewed, and has two distinct modes in designs I, II, and III, respectively. Laplace (ordinary smooth) and normal (super-smooth) distributions of ε_{it} are considered. For each design, the constant σ_0 is chosen so that $V[m(X_{it}, \alpha_i)] = V[U_{it}]$, i.e. so that the signal-to-noise ratio is equal to one. The function $m(x, \alpha)$ is chosen so that it is almost flat for small values of α , but is pronouncedly cubic when α is large; see also Figure 1.

Each Monte Carlo experiment is based on 500 repetitions. To avoid boundary effects, all estimators are calculated on the grid of points $x \in [0.1, 0.9]$, but boundary kernels are not used. Also, the rearranged conditional deconvolution CDF estimators $\tilde{F}_{m(x_t, \alpha_i)}(w|x_t)$ and $\tilde{F}_{m(x_t, \alpha_i)}(w|x_{t\tau})$ are further adjusted so that their lowest value is made to be zero, while the highest value is made to be one. More precisely, the estimator of the CDF $F_{m(x_t, \alpha_i)}(w|x_t)$ is obtained using the following procedure:

1. On the interval $[\hat{Q}_{Y_{it}}(0.01|X_{it} = x_t), \hat{Q}_{Y_{it}}(0.99|X_{it} = x_t)]$ a grid $G = \{\omega_1, \dots, \omega_{100}\}$ of 100 equally spaced points is generated, where $\hat{Q}_{Y_{it}}(q|X_{it} = x_t)$ is the conditional quantile estimator of Bhattacharya and Gangopadhyay (1990);

³¹ $\chi^2(r)$ with $r \leq 3$ have nonzero or nondifferentiable density on the boundary of the support, and hence do not satisfy the assumptions of the previous section.

2. $\hat{F}_{m(x_t, \alpha_i)}(\omega_j|x_t)$ is calculated for each j and its rearranged version $\tilde{F}_{m(x_t, \alpha_i)}(\omega_j|x_t)$ is obtained as discussed in Section 3.1;³²
3. The range corrected estimator $\hat{F}_{m(x_t, \alpha_i)}^{corr}(\omega_j|x_t)$ is obtained as

$$\hat{F}_{m(x_t, \alpha_i)}^{corr}(\omega_j|x_t) = \frac{\tilde{F}_{m(x_t, \alpha_i)}(\omega_j|x_t) - \tilde{F}_{m(x_t, \alpha_i)}(\omega_1|x_t)}{\tilde{F}_{m(x_t, \alpha_i)}(\omega_{100}|x_t) - \tilde{F}_{m(x_t, \alpha_i)}(\omega_1|x_t)}.$$

The resulting estimator $\hat{F}_{m(x_t, \alpha_i)}^{corr}(\omega|x_t)$ is used to calculate the estimator $\hat{Q}_{m(x_t, \alpha_i)}(\omega|x_t)$ (by numerical inversion of $\hat{F}_{m(x_t, \alpha_i)}^{corr}(\omega|x_t)$) and the estimators $\hat{m}_{FE}(x, \alpha)$ and $\hat{h}(x, q)$ (as described in Section 3.1). Note that the second step of this procedure (the rearrangement step) is shown to improve the estimates of CDFs in finite samples by Chernozhukov, Fernandez-Val, and Galichon (2007). The third step (range adjusting) seems to be useful in practice, though it may lack sound theoretical justification. This step is performed because the range of the estimated CDFs $\hat{F}_m(\omega_j|x_t)$ and $\tilde{F}_m(\omega_j|x_t)$ is often considerably different from $[0, 1]$, and hence some finite sample adjustment is desirable.

The values $h_w \in \{0.2, 0.4, 0.6\}$ are considered for deconvolution bandwidths. Bandwidths $h_Y(1)$, $h_Y(2)$, and h_U are all taken to be equal and are denoted by h_Y (this is clearly a suboptimal choice, since optimal $h_Y(1)$ should be taken larger than $h_Y(2)$ and h_U). The values $h_Y \in \{0.2, 0.4, 0.6\}$ are considered.

Tables 1-6 present the results of the Monte Carlo experiment. The presented Root Integrated Mean Squared Error (RIMSE), Root Integrated Squared Bias (RIBIAS²), and Root Integrated Variance (RIVAR) values are calculated as

$$\begin{aligned} RIMSE(q) &= \sqrt{\frac{1}{17} \sum_{l=0}^{16} R^{-1} \sum_{r=1}^R \left(\hat{h}_r(x_l, q) - h(x_l, q) \right)^2}, \\ RIBIAS^2(q) &= \sqrt{\frac{1}{17} \sum_{l=0}^{16} \left(R^{-1} \sum_{r=1}^R \hat{h}_r(x_l, q) - h(x_l, q) \right)^2}, \\ RIVAR(q) &= \sqrt{\frac{1}{17} \sum_{l=0}^{16} \left(R^{-1} \sum_{r=1}^R \hat{h}_r^2(x_l, q) - \left(R^{-1} \sum_{r=1}^R \hat{h}_r(x_l, q) \right)^2 \right)}, \end{aligned}$$

where $x_l = 0.1 + 0.05l$ is the l -th point of the grid over x , $\hat{h}_r(x, q)$ is the estimate of the true function $h(x, q)$ obtained in the r -th Monte Carlo replication, $R = 500$ is the number of replications, and $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$.

³²That is, $\tilde{F}_m(\omega_1|x_t)$ is taken to be the smallest of the values $\{\hat{F}_m(\omega_1|x_t), \dots, \hat{F}_m(\omega_{100}|x_t)\}$, $\tilde{F}_m(\omega_2|x_t)$ is taken to be the second smallest value and so on.

To give an idea of the properties of the estimation method, the following infeasible estimator is also simulated. First, the values of

$$\hat{h}_{INF}(x, q) = \frac{\sum_{t=1}^2 \sum_{i=1}^n Y_{it} K\left(\frac{X_{it}-x}{\sigma_{X_{it}} h^*}\right) K\left(\frac{\alpha_i - Q_{\alpha_i}(q)}{\sigma_{\alpha_i} h^*}\right)}{\sum_{t=1}^2 \sum_{i=1}^n K\left(\frac{X_{it}-x}{\sigma_{X_{it}} h^*}\right) K\left(\frac{\alpha_i - Q_{\alpha_i}(q)}{\sigma_{\alpha_i} h^*}\right)},$$

are calculated, where $Q_{\alpha_i}(q)$, $\sigma_{X_{it}}$, and σ_{α_i} are population quantities. This estimator is a kernel regression that uses the data on α_i , i.e. it estimates model (1) *as if* α_i were observed.³³ Moreover, this estimator and the corresponding RIMSE is calculated for a range of different values of the bandwidth h^* and then the best of these values of RIMSE is reported in the tables of results (thus this estimator also uses an infeasible bandwidth, which optimizes its finite sample behavior). It is of interest to see how the feasible estimator $\hat{h}(x, q)$ compares to this infeasible estimator.

Several observations are to be made regarding Table 1. First, it is clear that the value of deconvolution bandwidth $h_w = 0.2$ is too small. Deconvolution is an ill-posed inverse problem and hence some regularization is needed. Deconvolution bandwidth h_w is too low, it provides too little regularization and therefore the estimator performs very badly. However, the deconvolution bandwidths $h_w \in \{0.4, 0.6\}$ yield far better results. Note also the model with normal disturbances U_{it} is harder to estimate than the model with Laplace disturbances, but not much harder.

Most importantly, the estimator $\hat{h}(\cdot)$ performs very well when compared to the infeasible estimator. For most quantiles q and bandwidths $(h_Y, h_w) \in \{0.4, 0.6\} \times \{0.4, 0.6\}$, the RIMSE of the feasible estimator is only twice larger than that of the infeasible estimator, which uses the ideal bandwidth and the data on the unobservable α_i .

These findings also hold for the larger sample size $n = 10000$ and other designs, see Tables 2-6.³⁴ As an illustration, the results of the Monte Carlo experiment for Design I with $n = 2500$ and $h_Y = h_w = 0.4$ are also presented graphically in Figure 1.

The finding that the proposed estimator performs well is not inconsistent with the previous studies in the statistical literature, many of which suggest poor performance of deconvolution estimators in finite samples. Most previous studies considered density deconvolution and used significantly smaller samples than in the experiments above. It is important that the estimator $\hat{h}(x, q)$ only relies on the estimation of the cumulative distribution functions, which are smoother and easier to estimate than the corresponding density functions. Also, the presented Monte Carlo experiments uses larger (although typical for micro-econometric studies) sample sizes.

³³In fact, for each x_l , the values of $\hat{h}_{INF}(x_l, q)$ were calculated on a grid of quantiles and then rearranged, to improve the finite sample performance of the infeasible estimator.

³⁴Since the bandwidths $h_Y = 0.2$ and $h_w = 0.2$ appear to be too small, the corresponding results are not presented in Tables 3-6.

5 Conclusion

This paper has considered a nonparametric panel data model with nonseparable individual-specific effects. Nonparametric identification and estimation require data on only two time periods. This paper derives the rates of convergence of the proposed estimators and presents a Monte Carlo study, which suggests that the estimators perform very well in finite samples.

Several extensions of model (1) are presented in Evdokimov (2009). That paper relaxes the assumption of additivity of the disturbance U_{it} in (1) by considering a nonparametric panel transformation model with unobserved heterogeneity. Evdokimov (2009) also relaxes the assumption of scalar persistent heterogeneity α_i by considering time varying $\alpha_{it} = W'_{it}\beta_i$, where β_i is a vector of individual specific coefficients and W_{it} are observed time varying covariates.

Evdokimov (2010) applies the methods of this paper to estimation of the union wage premium. The union wage premium is found to be a decreasing function of the unobserved skill, especially for the above median skill levels. Nonparametric i.i.d. bootstrap is used to construct confidence intervals for the union wage premium function, although a formal proof of consistency of the bootstrap procedure is left for future work.

6 Appendix

To simplify the notation below, subscript m stands for $m(x_t, \alpha_i)$. Thus, $F_m(\omega|x_{t\tau})$ means $F_{m(\alpha_i, x_i)|X_{it}, X_{i\tau}}(\omega|x_t, x_\tau)$.

6.1 Proof of Lemma 1

Note that due to (i),

$$\begin{aligned} \phi_{(Y_1, Y_2)}(s_1, s_2) &= E[\exp(i(s_1 + s_2)A + is_1U_1 + is_2U_2)] \\ &= \phi_A(s_1 + s_2)\phi_{U_1}(s_1)\phi_{U_2}(s_2), \\ \frac{\partial \phi_{(Y_1, Y_2)}(s_1, s_2)}{\partial s_1} &= \phi'_A(s_1 + s_2)\phi_{U_1}(s_1)\phi_{U_2}(s_2) \\ &\quad + \phi_A(s_1 + s_2)\phi'_{U_1}(s_1)\phi_{U_2}(s_2), \end{aligned}$$

where the existence of the derivatives of characteristic function here and below is guaranteed by (ii) and the dominated convergence theorem. Then, using (iv) for all $s \in \mathbb{R}$ we obtain

$$\frac{\partial \phi_{(Y_1, Y_2)}(s, -s) / \partial s_1}{\phi_{(Y_1, Y_2)}(s, -s)} = \frac{\phi'_A(0)}{\phi_A(0)} + \frac{\phi'_{U_1}(s)}{\phi_{U_1}(s)},$$

where $\phi_A(0) = 1$ and $\phi'_A(0) = iE[A] = iE[Y_1]$ due to (iii). Therefore, for any $s \in \mathbb{R}$,

$$\begin{aligned}\phi_{U_1}(s) &= \exp\left(\int_0^s \frac{\partial \phi_{(Y_1, Y_2)}(\xi, -\xi) / \partial s_1}{\phi_{(Y_1, Y_2)}(\xi, -\xi)} d\xi - isE[Y_1]\right) \\ &= \exp\left(\int_0^s \frac{iE[Y_1 \exp(i\xi(Y_1 - Y_2))]}{\phi_{Y_1 - Y_2}(\xi)} d\xi - isE[Y_1]\right).\end{aligned}\quad (\text{A.1})$$

Using (iv),

$$\phi_A(s) = \frac{\phi_{(Y_1, Y_2)}(s, 0)}{\phi_{U_1}(s)} = \frac{\phi_{Y_1}(s)}{\phi_{U_1}(s)}, \quad \phi_{U_2}(s) = \frac{\phi_{(Y_1, Y_2)}(-s, s)}{\phi_{U_1}(-s)} = \frac{\phi_{Y_2 - Y_1}(s)}{\phi_{U_1}(-s)}. \quad \blacksquare$$

6.2 Identification with Serially Correlated Disturbances

Now suppose that the disturbance U_{it} in model (1) follows an autoregressive process of order one (AR(1)), i.e. $U_{it} = \rho U_{it-1} + \varepsilon_{it}$ for some constant ρ , $|\rho| < 1$. Then, the model can be identified using a panel with three or more time periods. Consider the following modifications of Assumption ID(i)-(vi):

Assumption AR. Suppose $U_{it} = \rho U_{it-1} + \varepsilon_{it}$ for all $t \geq 2$ and $|\rho| < 1$. Also, suppose that:

- (i) $\{X_i, U_i, \alpha_i\}_{i=1}^n$ is a random sample and $T = 3$;
- (ii) $f_{\varepsilon_{it}|X_{it}, \alpha_i, X_{i(-t)}, \varepsilon_{i(-t)}, U_{i1}}(\varepsilon_t | x_t, \alpha, x_{(-t)}, \varepsilon_{(-t)}, u_1) = f_{\varepsilon|X_{it}}(\varepsilon_t | x_t)$ and for all $t \in \{2, 3\}$ and $(\varepsilon_t, x_t, \alpha, x_{(-t)}, \varepsilon_{(-t)}) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X}^2 \times \mathbb{R}$,³⁵
- (iib) $f_{U_{i1}|X_{i1}, \alpha_i, X_{i(-1)}, \varepsilon_i}(u_1 | x_1, \alpha, x_{(-1)}, \varepsilon) = f_{U_{i1}|X_{i1}}(u_1 | x_1)$ for all $(u_1, x_1, \alpha, x_{(-1)}, \varepsilon) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X}^2 \times \mathbb{R}^2$;
- (iii) $E[U_{i1}|X_i = (x_1, x_2, x_3)] = 0$ and $E[\varepsilon_{it}|X_i = (x_1, x_2, x_3)] = 0$ for all $(x_1, x_2, x_3) \in \mathcal{X}^3$ and $t \geq 2$, where $X_i = (X_{i1}, X_{i2}, X_{i3})$;
- (iv) the conditional characteristic functions $\phi_{U_{i1}}(s|X_{i1} = x)$, $\phi_{\varepsilon_{it}}(s|X_{it} = x)$ do not vanish for all $s \in \mathbb{R}$, $x \in \mathcal{X}$, and $t \in \{2, 3\}$;
- (v) $E[|m(x_t, \alpha_i)| + |U_{i1}| + |\varepsilon_{i2}| + |\varepsilon_{i3}| | X_i = (x_1, x_2, x_3)]$ is uniformly bounded for all t and $(x_1, x_2, x_3) \in \mathcal{X}^3$;
- (vi) for each $x \in \mathcal{X}$ there is a $x_1(x) \in \mathcal{X}$ such that $f_{X_{i1}, X_{i2}, X_{i3}}(x_1(x), x, x) > 0$; also, $f_{X_{i1}, X_{i2}}(x, x) > 0$ for all $x \in \mathcal{X}$.

Assumptions AR(ii)-(iib) are similar to Assumption ID(ii) and imply that $\alpha_i, \varepsilon_{it}, \varepsilon_{is}, U_{i1}$ are mutually independent for all $t, s \geq 2, t \neq s$. However, there is an important issue regarding Assumption AR(iib) on U_{i1} . For illustration, suppose that $\varepsilon_{it} = \sigma_t(X_{it})\xi_{it}$ for all t , where $\xi_{it} \sim i.i.d. (0, 1)$ are independent of α_i and $\{X_{ij}\}_{j=-\infty}^{\infty}$. Assuming that the data have infinite history, $U_{i1} = \sum_{j=0}^{\infty} \rho^j \sigma_{1-j}(X_{i,1-j})\varepsilon_{i,1-j}$. Then, it is hard to guarantee $\alpha_i \perp U_{i1}$, conditional

³⁵Here index $(-t)$ stands for "other than t " time periods. Note that $x_{(-t)}$ belongs to \mathcal{X}^2 , while $\varepsilon_{(-t)}$ is a scalar when $T = 3$ and $t \geq 2$ because there is no ε_{i1} in the model.

on $X_i = (X_{i1}, X_{i2}, X_{i3})$. The reason is that α_i may be correlated with the past covariates $\{X_{it}\}_{t \leq 0}$, which affect the variance of past shocks $\{\varepsilon_{it}\}_{t \leq 0}$ and hence the distribution of U_{i1} . This is the initial condition problem. There are at least two ways to ensure that the conditional independence condition $\alpha_i \perp U_{i1} | X_i$ holds in this model. First, one can assume that $\sigma_t(X_{it}) = \sigma_t$, i.e. that the innovations ε_{it} are independent of the covariates X_{it} . Alternatively, in some cases there is no past history of ε_{it} , and U_{i1} is a true initialization condition. For instance, this assumption may hold when a firm is present in the data since the date it was established.

Note that Assumption AR(vi) is a very weak extension of assumption ID(vi). Also, note that Assumption AR does not require ε_{it} to have the same conditional distribution for different $t \in \{2, 3\}$.

Theorem 9. *Suppose Assumptions AR, ID(vii)-(viii), and either RE or FE are satisfied. Then model (1) is identified.*

Proof. 1. Take any $x \in \mathcal{X}$ and note that $E[(Y_{i3} - Y_{i2})Y_{i1} | X_i = (x_1(x), x, x)] = 0$ iff $\rho = 0$. When $\rho = 0$ identification of $\phi_{U_{i1}}(s | X_{i1} = x)$ and $\phi_{\varepsilon_{it}}(s | X_{it} = x)$ follows immediately from the first step of Theorem 1. When $\rho \neq 0$ using Assumption AR(ii-iii) one obtains

$$\rho = E[(Y_{i3} - Y_{i2})Y_{i1} | X_i = (x_1(x), x, x)] / E[(Y_{i2} - Y_{i1})Y_{i1} | X_{i1} = X_{i2} = x_1(x)].$$

Then for any $x \in \mathcal{X}$

$$\begin{aligned} & \phi_{Y_{i1}, Y_{i2}}(s_1, s_2 | X_{i1} = X_{i2} = x) \\ &= E[\exp\{i[(s_1 + s_2)m(x, \alpha_i) + (s_1 + \rho s_2)U_{i1} + s_2\varepsilon_{i2}]\} | X_{i1} = X_{i2} = x] \\ &= \phi_{m(x, \alpha_i)}(s_1 + s_2 | X_{i1} = X_{i2} = x) \phi_{U_{i1}}(s_1 + \rho s_2 | X_{i1} = x) \phi_{\varepsilon_{i2}}(s_2 | X_{i2} = x). \end{aligned} \quad (\text{A.2})$$

Denote $\mu(x) = E[m(x, \alpha_i) | X_{i1} = X_{i2} = x] = E[Y_{i1} | X_{i1} = X_{i2} = x]$. Then, similar to the proof of Lemma 1, for all $s \in \mathbb{R}$

$$\frac{\partial \phi_{Y_1, Y_2}(s, -s) / \partial s_1}{\phi_{Y_1, Y_2}(s, -s)} = i\mu(x) + \frac{\partial \phi_{U_{i1}}((1 - \rho)s | X_{i1} = x) / \partial s}{\phi_{U_{i1}}((1 - \rho)s | X_{i1} = x)}.$$

Thus, one obtains

$$\phi_{U_{i1}}(s | X_{i1} = x) = \exp\left(\int_0^s \frac{\partial \phi_{(Y_1, Y_2)}\left(\left(1 - \rho\right)^{-1} \xi, -\left(1 - \rho\right)^{-1} \xi\right) / \partial s_1}{\phi_{(Y_1, Y_2)}\left(\left(1 - \rho\right)^{-1} \xi, -\left(1 - \rho\right)^{-1} \xi\right)} d\xi - is\mu(x)\right), \quad (\text{A.3})$$

and using (A.2)

$$\phi_{\varepsilon_{i2}}(s | X_{i2} = x) = \frac{\phi_{Y_{i1}, Y_{i2}}(-s, s | X_{i1} = X_{i2} = x)}{\phi_{U_{i1}}(s(\rho - 1) | X_{i1} = x)}.$$

2. Now consider identification under Assumption FE. Note that for any $x \in \mathcal{X}$

$$\begin{aligned} \phi_{Y_{i1}}(s | X_{i1} = x, X_{i2} = \bar{x}) &= \phi_{m(x, \alpha_i)}(s | X_{i1} = x, X_{i2} = \bar{x}) \phi_{U_{i1}}(s | X_{i1} = x), \\ \phi_{Y_{i2}}(s | X_{i1} = x, X_{i2} = \bar{x}) &= \phi_{m(\bar{x}, \alpha_i)}(s | X_{i1} = x, X_{i2} = \bar{x}) \phi_{U_{i1}}(\rho s | X_{i1} = x) \phi_{\varepsilon_{i2}}(s | X_{i2} = \bar{x}), \end{aligned}$$

thus $\phi_{m(x, \alpha_i)}(s | X_{i1} = x, X_{i2} = \bar{x})$ and $\phi_{m(\bar{x}, \alpha_i)}(s | X_{i1} = x, X_{i2} = \bar{x})$ are identified. The rest

of the proof follows step 3 of the proof of Theorem 2. The proof of identification under Assumption RE is similar. ■

Now suppose that the disturbance term follows moving average process of order one (MA(1)). To demonstrate the flexibility of the framework, the MA(1) model is identified under somewhat different assumptions than the AR(1) model above. Assume that the disturbance term follows $U_{it} = \varepsilon_{it} + \theta(X_{it}) \varepsilon_{it-1}$, i.e. the moving average parameter depends on the value of the covariate in the corresponding time period.

Assumption MA. *Suppose $U_{it} = \varepsilon_{it} + \theta(X_{it}) \varepsilon_{it-1}$, for all $t \geq 1$, and $\sup_{x \in \mathcal{X}} |\theta(x)| < 1$. Suppose also that:*

- (i) $\{X_i, U_i, \alpha_i\}_{i=1}^n$ is a random sample and $T = 3$;
- (iia) $f_{\varepsilon_{it}|X_{it}, \alpha_i, X_{i(-t)}, \varepsilon_{i(-t)}}(\varepsilon_t | x_t, \alpha, x_{(-t)}, \varepsilon_{(-t)}, u_1) = f_{\varepsilon|X_{it}}(\varepsilon_t | x_t)$ and for all $(\varepsilon_t, x_t, \alpha, x_{(-t)}, \varepsilon_{(-t)}) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X}^2 \times \mathbb{R}$ and $t \geq 1$,³⁶
- (iib) $f_{\varepsilon_{i0}|X_{i1}, \alpha_i, X_{i(-1)}, \varepsilon_i}(\varepsilon_0 | x_1, \alpha, x_{(-1)}, \varepsilon) = f_{\varepsilon_{i0}|X_{i1}}(\varepsilon_0 | x_1)$ for all $(\varepsilon_0, x_1, \alpha, x_{(-1)}, \varepsilon) \in \mathbb{R} \times \mathcal{X} \times \mathbb{R} \times \mathcal{X}^2 \times \mathbb{R}^3$;
- (iii) $E[\varepsilon_{it} | X_i = (x_1, x_2, x_3)] = 0$ for all $(x_1, x_2, x_3) \in \mathcal{X}^3$ and $t \geq 0$;
- (iv) the conditional characteristic functions $\phi_{\varepsilon_{it}}(s | X_{it} = x)$ and $\phi_{\varepsilon_{i0}}(s | X_{i1} = x)$ do not vanish for all $s \in \mathbb{R}$, $x \in \mathcal{X}$, and $t \geq 1$;
- (v) $E[|m(x_t, \alpha_i)| + \sum_{j=0}^3 |\varepsilon_{ij}| | X_i = (x_1, x_2, x_3)]$ is uniformly bounded for all t and $(x_1, x_2, x_3) \in \mathcal{X}^3$;
- (vi) $f_{X_{i1}, X_{i2}, X_{i3}}(x, x, x) > 0$ for each $x \in \mathcal{X}$.

Assumptions MA(i-v) are similar to Assumptions AR(i-v). Note the "initial condition" type Assumption MA(iib). In contrast to Assumption AR(vi), Assumption MA(vi) requires that the trivariate density $f_{X_{i1}, X_{i2}, X_{i3}}(x, x, x)$ is non-zero. This is the price paid for allowing the moving average parameter θ to depend on X_{it} .

Theorem 10. *Suppose Assumptions MA, ID(vii)-(viii), and either RE or FE are satisfied. Then model (1) is identified.*

Proof. 1. Fix any $x \in \mathcal{X}$ and define the event $\mathcal{G}_x = \{X_i = (x, x, x)\}$ and the function $\sigma_t^2(x) = E[\varepsilon_{it}^2 | X_{it} = x]$ for $t \geq 1$. Also define

$$\begin{aligned} A_1(x) &= E[(Y_{i2} - Y_{i3}) Y_{i1} | \mathcal{G}_x] = \theta(x) \sigma_1^2(x) \\ A_2(x) &= E[(Y_{i2} - Y_{i1}) Y_{i3} | \mathcal{G}_x] = \theta(x) \sigma_2^2(x) \\ A_3(x) &= E[(Y_{i2} - Y_{i3}) Y_{i2} | \mathcal{G}_x] = \theta^2(x) \sigma_1^2(x) + (1 - \theta(x)) \sigma_2^2(x), \end{aligned}$$

Note that $\theta(x) = 0$ iff $A_1(x) = 0$. When $\theta(x)$ identification follows the first step of Theorem 1. When $\theta(x) \neq 0$ one can write $A_1(x) + A_3(x) = \theta^2(x) \sigma_1^2(x) + \sigma_2^2(x)$ and obtain an equation for $\theta(x)$:

$$A_1(x) + A_3(x) = \theta(x) A_1(x) + A_2(x) / \theta(x),$$

³⁶Here $\varepsilon_{(-t)} \in \mathbb{R}^3$ when $T = 3$. For instance, $\varepsilon_{(-2)} = (\varepsilon_0, \varepsilon_1, \varepsilon_3)$.

which has unique solution given that $|\theta(x)| < 1$. Thus $\theta(x)$ is identified.

Note that

$$\begin{aligned} \phi_{Y_1, Y_2, Y_3}(s_1, s_2, s_3 | \mathcal{G}_x) &= \phi_{m(x, \alpha_i)}(s_1 + s_2 + s_3 | \mathcal{G}_x) \phi_{\varepsilon_{i0}}(s_1 \theta(x) | X_{i1} = x) \times \\ &\phi_{\varepsilon_{i1}}(s_2 \theta(x) + s_1 | X_{i1} = x) \phi_{\varepsilon_{i2}}(s_3 \theta(x) + s_2 | X_{i2} = x) \phi_{\varepsilon_{i3}}(s_3 | X_{i3} = x). \end{aligned} \quad (\text{A.4})$$

Note that $\partial \phi_{\varepsilon_{i3}}(0 | X_{i3} = x) / \partial s_3 = 0$ due to Assumption AR(iii). Then, for all s

$$\frac{\partial \phi_{Y_1, Y_2, Y_3}(-s, s, 0) / \partial s_3}{\phi_{Y_1, Y_2, Y_3}(-s, s, 0)} = \mathbf{i}E[m(x, \alpha_i) | \mathcal{G}_x] + \theta(x) \frac{\partial \phi_{\varepsilon_{i2}}(s | X_{i2} = x) / \partial s}{\phi_{\varepsilon_{i2}}(s | X_{i2} = x)}.$$

Thus, $\phi_{\varepsilon_{i2}}(s | X_{i2} = x)$ is identified and an expression similar to (A.3) can be given for $\phi_{\varepsilon_{i2}}(s | X_{i2} = x)$. Similarly, $\phi_{\varepsilon_{i1}}(s | X_{i1} = x)$ is identified by the equation

$$\frac{\partial \phi_{Y_1, Y_2, Y_3}(0, s, -s) / \partial s_1}{\phi_{Y_1, Y_2, Y_3}(0, s, -s)} = \mathbf{i}E[m(x, \alpha_i) | \mathcal{G}_x] + \frac{\partial \phi_{\varepsilon_{i1}}(s \theta(x) | X_{i1} = x) / \partial s}{\phi_{\varepsilon_{i1}}(s \theta(x) | X_{i1} = x)}.$$

Then, functions $\phi_{\varepsilon_{i3}}(s | X_{i3} = x)$ and $\phi_{\varepsilon_{i0}}(s | X_{i1} = x)$ are identified, respectively, by equations

$$\begin{aligned} \frac{\partial \phi_{Y_1, Y_2, Y_3}(-s, 0, s) / \partial s_3}{\phi_{Y_1, Y_2, Y_3}(-s, 0, s)} &= \mathbf{i}E[m(x, \alpha_i) | \mathcal{G}_x] + \theta(x) \frac{\partial \phi_{\varepsilon_{i2}}(\theta(x) s | X_{i2} = x) / \partial s}{\phi_{\varepsilon_{i2}}(\theta(x) s | X_{i2} = x)} \\ &+ \frac{\partial \phi_{\varepsilon_{i3}}(s | X_{i3} = x) / \partial s}{\phi_{\varepsilon_{i3}}(s | X_{i3} = x)}, \text{ and} \end{aligned}$$

$$\begin{aligned} \frac{\partial \phi_{Y_1, Y_2, Y_3}(s, 0, -s) / \partial s_1}{\phi_{Y_1, Y_2, Y_3}(s, 0, -s)} &= \mathbf{i}E[m(x, \alpha_i) | \mathcal{G}_x] + \theta(x) \frac{\partial \phi_{\varepsilon_{i0}}(\theta(x) s | X_{i1} = x) / \partial s}{\phi_{\varepsilon_{i0}}(s | X_{i1} = x)} \\ &+ \frac{\partial \phi_{\varepsilon_{i1}}(s | X_{i1} = x) / \partial s}{\phi_{\varepsilon_{i1}}(s | X_{i1} = x)}. \end{aligned}$$

The rest of the proof is similar to the second step of the proof of Theorem 9

2. Suppose Assumption FE holds. For any $x \in \mathcal{X}$

$$\begin{aligned} \phi_{Y_{i1}}(s | X_{i1} = x, X_{i2} = \bar{x}) &= \phi_{m(x, \alpha_i)}(s | X_{i1} = x, X_{i2} = \bar{x}) \phi_{\varepsilon_{i1}}(s | X_{i1} = x) \phi_{\varepsilon_{i0}}(\theta(x) s | X_{i1} = x), \\ \phi_{Y_{i2}}(s | X_{i1} = x, X_{i2} = \bar{x}) &= \phi_{m(\bar{x}, \alpha_i)}(s | X_{i1} = x, X_{i2} = \bar{x}) \phi_{\varepsilon_{i2}}(s | X_{i2} = \bar{x}) \phi_{\varepsilon_{i1}}(\theta(\bar{x}) s | X_{i1} = x), \end{aligned}$$

and hence functions $\phi_{m(x, \alpha_i)}(\cdot)$ and $\phi_{m(\bar{x}, \alpha_i)}(\cdot)$ are identified and the rest of the proof follows step 3 of the proof of Theorem 2. The proof of identification under Assumption RE is similar. \blacksquare

It is straightforward to extend the approach to larger ARMA(p,q) models.

6.3 Proof of Identification When X_{it} Has Continuous Components

Define the events $\mathcal{G}_{X_{it}, n}(x) = \{X_{it} \in (x - 1_p/n, x + 1_p/n)\}$ and $\mathcal{G}_n(x) = \mathcal{G}_{X_{i1}, n}(x) \cap \mathcal{G}_{X_{i2}, n}(x)$, where 1_p is the identity vector of length $p = \dim(\mathcal{X})$. Fix any $x_0 \in \mathcal{X}$ and $s \in \mathbb{R}$. For any

$n > 0$ we have

$$\begin{aligned} \frac{\partial \phi_{Y_{i1}, Y_{i2}}(s, -s | \mathcal{G}_n(x_0)) / \partial s}{\phi_{Y_{i1}, Y_{i2}}(s, -s | \mathcal{G}_n(x_0))} &= \frac{E[\text{im}(X_{i1}, \alpha_i) \exp\{\text{is}(m(X_{i1}, \alpha_i) - m(X_{i2}, \alpha_i))\} | \mathcal{G}_n(x_0)]}{E[\exp\{\text{is}(m(X_{i1}, \alpha_i) - m(X_{i2}, \alpha_i))\} | \mathcal{G}_n(x_0)]} \\ &\quad + \frac{\partial \phi_{U_{i1}}(s | \mathcal{G}_{X_{i1}, n}(x_0)) / \partial s}{\phi_{U_{i1}}(s | \mathcal{G}_{X_{i1}, n}(x_0))}. \end{aligned}$$

First, for any $\delta > 0$ there exists a large enough $n_1(\delta)$ such that $\sup_{(x_1, x_2) \in (x_0 - 1_p/n, x_0 + 1_p/n)^2} \sup_{\alpha \in [-\delta^{-1/2}, \delta^{-1/2}]} |m(x_1, \alpha) - m(x_2, \alpha)| \leq \delta$ holds for all $n \geq n_1(\delta)$. Then, for any $n \geq n_1(\delta)$:

$$\begin{aligned} &|E[\text{im}(X_{i1}, \alpha_i) \exp\{\text{is}(m(X_{i1}, \alpha_i) - m(X_{i2}, \alpha_i))\} | \mathcal{G}_n(x_0)] - E[\text{im}(X_{i1}, \alpha_i) | \mathcal{G}_n(x_0)]| \\ &= \left| \int_{(x_1, x_2) \in (x_0 - 1_p/n, x_0 + 1_p/n)^2} \int_{\alpha \in \mathbb{R}} m(x_1, \alpha) (\exp\{\text{is}(m(x_1, \alpha) - m(x_2, \alpha))\} - 1) \right. \\ &\quad \left. \times f_{\alpha_i}(\alpha | x_1, x_2) d\alpha dF_{X_{i1}, X_{i2}}(x_1, x_2 | \mathcal{G}_n(x_0)) \right| \\ &\leq 2 \int_{(x_1, x_2) \in (x_0 - 1_p/n, x_0 + 1_p/n)^2} \int_{|\alpha| > \delta^{-1/2}} |m(x_1, \alpha)| f_{\alpha_i}(\alpha | x_1, x_2) d\alpha dF_{X_{i1}, X_{i2}}(x_1, x_2 | \mathcal{G}_n(x_0)) \\ &\quad + \left| \int_{(x_1, x_2) \in (x_0 - 1_p/n, x_0 + 1_p/n)^2} \int_{|\alpha| \leq \delta^{-1/2}} \right. \\ &\quad \left. m(x_1, \alpha) \int_0^s \text{i}(m(x_1, \alpha) - m(x_2, \alpha)) \exp\{\text{i}\xi(m(x_1, \alpha) - m(x_2, \alpha))\} d\xi \right. \\ &\quad \left. \times f_{\alpha_i | X_{i1}, X_{i2}}(\alpha | x_1, x_2) d\alpha dF_{X_{i1}, X_{i2}}(x_1, x_2 | \mathcal{G}_n(x_0)) \right| \\ &\leq 2E\left[|m(X_{i1}, \alpha_i)| \mathbf{1}\{|\alpha_i| > \delta^{-1/2}\} \middle| \mathcal{G}_n(x_0)\right] \\ &\quad + \delta |s| E\left[\left[m(X_{i1}, \alpha_i) \mathbf{1}\{|\alpha_i| \leq \delta^{-1/2}\} \middle| \mathcal{G}_n(x_0)\right]\right] \\ &\rightarrow 0 \text{ as } \delta \rightarrow 0. \end{aligned}$$

Second, for any $\delta > 0$ there exists a large enough $n_2(\delta)$ such that

$$\sup_{(x_1, x_2, \alpha) \in (x_0 - 1_p/n, x_0 + 1_p/n)^2 \times [-\delta^{-1/2}, \delta^{-1/2}]} |m(x_1, \alpha) f_{\alpha_i}(\alpha | x_1, x_2) - m(x_0, \alpha) f_{\alpha_i}(\alpha | x_0, x_0)| \leq \delta$$

holds for all $n \geq n_2(\delta)$. Then, for all $n \geq n_2(\delta)$:

$$\begin{aligned}
& |E[m(X_{i1}, \alpha_i) | \mathcal{G}_n(x_0)] - E[m(X_{i1}, \alpha_i) | X_{i1} = X_{i2} = x_0]| \\
= & \left| \int_{\alpha \in \mathbb{R}} \int_{(x_1, x_2) \in (x_0 - 1_p/n, x_0 + 1_p/n)^2} m(x_1, \alpha) f_{\alpha_i}(\alpha | x_1, x_2) dF_{X_{i1}, X_{i2}}(x_1, x_2 | \mathcal{G}_{X_{it}, n}(x_0)) d\alpha \right. \\
& \left. - \int_{\alpha \in \mathbb{R}} m(x_0, \alpha) f_{\alpha_i}(\alpha | x_0, x_0) d\alpha \right| \\
\leq & 2\delta^{1/2} + E \left[|m(X_{i1}, \alpha_i)| 1 \left\{ |\alpha_i| > \delta^{-1/2} \right\} \middle| \mathcal{G}_n(x_0) \right] \\
& + E \left[|m(X_{i1}, \alpha_i)| 1 \left\{ |\alpha_i| > \delta^{-1/2} \right\} \middle| X_{i1} = X_{i2} = x_0 \right] \\
\rightarrow & 0 \text{ as } \delta \rightarrow 0,
\end{aligned}$$

where the inequality follows from splitting the integration over the regions of $|\alpha| \geq \delta^{-1/2}$ as before and the triangle inequality. Taking $n_3(\delta) = \max\{n_1(\delta), n_2(\delta)\}$ we get that for all $n \geq n_3(\delta)$

$$\begin{aligned}
& |E[im(X_{i1}, \alpha_i) \exp\{is(m(X_{i1}, \alpha_i) - m(X_{i2}, \alpha_i))\} | \mathcal{G}_n(x_0)] \\
& \qquad \qquad \qquad - E[m(X_{i1}, \alpha_i) | X_{i1} = X_{i2} = x_0]| \\
\rightarrow & 0 \text{ as } \delta \rightarrow 0.
\end{aligned}$$

Similarly, one can show that

$$\begin{aligned}
E[\exp\{is(m(X_{i1}, \alpha_i) - m(X_{i2}, \alpha_i))\} | \mathcal{G}_n(x_0)] & \rightarrow 1, \\
\phi_{U_{i1}}(s | \mathcal{G}_{X_{i1}, n}(x_0)) & \rightarrow \phi_{U_{i1}}(s | X_{i1} = x_0), \\
\partial \phi_{U_{i1}}(s | \mathcal{G}_{X_{i1}, n}(x_0)) / \partial s = E[iU_{i1} \exp\{isU_{i1}\} | \mathcal{G}_n(x_0)] & \rightarrow \partial \phi_{U_{i1}}(s | X_{i1} = x_0) / \partial s,
\end{aligned}$$

as $n \rightarrow \infty$.

Thus, as $n \rightarrow \infty$:

$$\frac{\partial \phi_{Y_{i1}, Y_{i2}}(s, -s | \mathcal{G}_n(x_0)) / \partial s}{\phi_{Y_{i1}, Y_{i2}}(s, -s | \mathcal{G}_n(x_0))} \rightarrow E[ism(X_{i1}, \alpha_i) | X_{i1} = X_{i2} = x_0] + \frac{\partial \phi_{U_{i1}}(s | X_{i1} = x_0) / \partial s}{\phi_{U_{i1}}(s | X_{i1} = x_0)},$$

where $E[ism(X_{i1}, \alpha_i) | X_{i1} = X_{i2} = x_0] = isE[isY_1 | X_{i1} = X_{i2} = x_0]$ is identified. Hence $\phi_{U_{i1}}(s | X_{i1} = x_0)$ is identified. Then, one identifies $\phi_{m(x_1, \alpha_i)}(s | X_{i1} = x_1, X_{i2} = x_2)$ and $\phi_{U_{i2}}(s | X_{i2} = x_0)$ similarly. ■

6.4 Identification with Misclassification

This section explains how the identification results of Theorems 1 and 2 can be extended to the analysis of data with misclassification when the probability of misclassification is either known or can be estimated from a separate dataset. To simplify the presentation of the main idea assume that the covariate is a scalar binary variable $X_{it}^* \in \{0, 1\}$. Suppose that the true

value of covariate X_{it}^* is not observed and instead one observes

$$X_{it} = \begin{cases} X_{it}^*, & \text{with probability } p, \\ 1 - X_{it}^*, & \text{with probability } 1 - p. \end{cases} \quad (\text{A.5})$$

Thus, X_{it} is a measurement of X_{it}^* that is subject to misclassification. Suppose also, that the events of misclassification are independent over time. An empirical example of such settings is given in the union membership study of Card (1996), where X_{it}^* is the union membership that is likely to be misreported. Card (1996) uses Current Population Survey data with a validation dataset and finds that the misclassification model (A.5) describes the data well. The validation dataset is also used to estimate the probability of misclassification p .

When the value of p is known (eg. from a validation study) we can write the joint characteristic function of (Y_{i1}, Y_{i2}) conditional on the observables X_{it} as

$$\begin{aligned} \phi_{Y_{i1}, Y_{i2} | X_{i1}, X_{i2}}(s_1, s_2 | 0, 0) &= p^2 \phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(s_1, s_2 | 0, 0) \\ &+ p(1-p) \phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(s_1, s_2 | 0, 1) \\ &+ p(1-p) \phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(s_1, s_2 | 1, 0) \\ &+ (1-p)^2 \phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(s_1, s_2 | 1, 1), \end{aligned}$$

where $\phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(\cdot)$ is the joint characteristic function of (Y_{i1}, Y_{i2}) conditional on the unobservables X_{it}^* . One can write similar representations of $\phi_{Y_{i1}, Y_{i2} | X_{i1}, X_{i2}}(s_1, s_2 | x_1, x_2)$ for $(x_1^*, x_2^*) \in \{(0, 1), (1, 0), (1, 1)\}$.

Then, for any (s_1, s_2) one obtains a system of four linear equations with four unknowns $\phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(s_1, s_2 | x_1^*, x_2^*)$. This system of equations has a unique solution if the misclassification probability p is not equal to $1/2$. Thus, one can identify $\phi_{Y_{i1}, Y_{i2} | X_{i1}^*, X_{i2}^*}(s_1, s_2 | x_1^*, x_2^*)$ and then follow the proofs of Theorems 1 and 2 to identify $m(x^*, \alpha)$ and other objects of interest.

6.5 Proofs of the Results on Estimation

Proof of Lemma 2. For simplicity of notation the proof is given for $\mathcal{X} \subset \mathbb{R}$. Generalization to a multivariate x is immediate. Note that

$$\begin{aligned} &\left| \frac{\partial^k [\partial \phi_{U_{it}}(s | x_t) / \partial s]}{\partial x_t^k} \right| = \left| \frac{\partial^k}{\partial x_t^k} \int_{-\infty}^{\infty} i u e^{isu} f_{U_{it}}(u | x_t) du \right| \\ &= \left| \int_{-\infty}^{\infty} i u e^{isu} \frac{\partial^k f_{U_{it}}(u | x_t) / \partial x_t^k}{f_{U_{it}}(u | x_t)} f_{U_{it}}(u | x_t) du \right| \\ &\leq \left(\int_{-\infty}^{\infty} u^2 f_{U_{it}}(u | x_t) du \cdot \int_{-\infty}^{\infty} \frac{(\partial^k f_{U_{it}}(u | x_t) / \partial x_t^k)^2}{f_{U_{it}}(u | x_t)} du \right)^{1/2} \leq C, \end{aligned}$$

where the second equality follows by the support assumption and continuity of $\partial^k f_{U_{it}}(u | x_t) / \partial x_t^k$, and the first inequality follows from Cauchy-Schwarz inequality. Thus, the k -th derivative of the function $\partial \phi_{U_{it}}(s | x_t) / \partial s$ is uniformly bounded for all s , hence $\partial \phi_{U_{it}}(s | x_t) / \partial s \in \mathcal{D}_{\mathcal{X}}^{\mathbb{R}}(k)$ is proved. The condition $\phi_{U_{it}}(s | x_t) \in \mathcal{D}_{\mathcal{X}}^{\mathbb{R}}(k)$ can be shown to hold in exactly the same way. ■

Proof of Lemma 3. Again, for simplicity of notation the proof is given for $\mathcal{X} \subset \mathbb{R}$. Similar

to the proof of Lemma 2:

$$\begin{aligned}
& \left| \frac{\partial^k \left[\partial \phi_{m(x_t, \alpha_i)}(s|x_t) / \partial s \right]}{\partial x_t^k} \right| = \left| \frac{\partial^k}{\partial x_t^k} \int_{\psi_l(x_t)}^{\psi_h(x_t)} i\omega(x_t, \alpha) f_{\alpha_i}(\alpha|x_t) d\alpha \right| \\
& = \left| \sum_{q=0}^k \binom{k}{q} \int_{\psi_l(x_t)}^{\psi_h(x_t)} \frac{\partial^q \omega(x_t, \alpha)}{\partial x_t^q} \frac{\partial^{k-q} f_{\alpha_i}(\alpha|x_t) / \partial x_t^{k-q}}{f_{\alpha_i}(\alpha|x_t)} f_{\alpha_i}(\alpha|x_t) du \right| \\
& \leq \sum_{q=0}^k \binom{k}{q} \left(\int_{\psi_l(x_t)}^{\psi_h(x_t)} \left| \frac{\partial^q \omega(x_t, \alpha)}{\partial x_t^q} \right|^2 f_{\alpha_i}(\alpha|x_t) d\alpha \cdot \int_{\psi_l(x_t)}^{\psi_h(x_t)} \frac{\left(\partial^{k-q} f_{\alpha_i}(\alpha|x_t) / \partial x_t^{k-q} \right)^2}{f_{\alpha_i}(\alpha|x_t)} d\alpha \right)^{1/2} \\
& \leq C \left(1 + s^k \right),
\end{aligned}$$

where the second equality follows by the support assumption and continuity of $\partial^k \omega(x_t, \alpha) / \partial x_t^k$ and $\partial^k f_{\alpha_i}(\alpha|x_t) / \partial x_t^k$, and the first inequality follows from Cauchy-Schwarz inequality. Hence, $\partial \phi_{m(x_t, \alpha_i)}(s|x_t) / \partial s \in \mathcal{D}_{\mathcal{X}}^{\mathbb{R}}(k)$ is proved. The other conditions of Assumption 5 can be shown to hold in a similar way. ■

Lemmas 4-6 are used in the proofs of Theorem 3:

Lemma 4. *Suppose Assumptions SYM, ID(i)-(vi), and 1-7 hold. Assume that $nh_U \rightarrow \infty$, $h_U \rightarrow 0$, $[\log(n)]^{1/2} n^{1/\gamma-1/2} h_U^{-p} \rightarrow 0$, $M_n \rightarrow \infty$, and $M_n h_U^p (\log(n))^{\gamma-3/2} n^{1/\gamma-1/2} \rightarrow 0$. Then,*

$$\begin{aligned}
& \sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \left| s^{-1} \left(\hat{\phi}_U^S(s|x) - \phi_U(s|x) \right) \right| \\
& = O_p \left(\chi(M_n) \left[\left[\log(n) / (nh_U^{2p}) \right]^{1/2} + \left(1 + M_n^{\bar{k}-1} \right) h_U^{\bar{k}} \right] \right).
\end{aligned}$$

Proof. 1. Let us first introduce some notation and then explain the logic of the proof. Consider any t and τ , $t < \tau$. Define the event $\mathcal{G}_x^{t\tau} = \{X_{it} = X_{i\tau} = x\}$. Define $\kappa_i^{t,\tau}(x, h) = K((X_{it} - x)/h) K((X_{i\tau} - x)/h)$ and $R_{i,t\tau}(x) = h_U^{-2p} \kappa_i^{t,\tau}(x, h_U)$. Define $Z_{i,t\tau} = Y_{it} - Y_{i\tau}$ and

$$\Delta_{n,t\tau}(s, x) = n^{-1} \sum_{i=1}^n s^{-1} \left(e^{isZ_{i,t\tau}} - E \left[e^{isZ_{i,t\tau}} | \mathcal{G}_x^{t\tau} \right] \right) R_{i,t\tau}(x).$$

By continuity, one can take $\Delta_{n,t\tau}(0, x) = n^{-1} \sum_{i=1}^n i \left[(Z_{i,t\tau} - E[Z_{i,t\tau} | \mathcal{G}_x^{t\tau}]) R_{i,t\tau}(x) \right]$, although the value of $\Delta_{n,t\tau}(s, x)$ at $s = 0$ is not important, because we are ultimately interested in approximating an integral over s of this function. In parts 1-3 of this proof t and τ are fixed, and to simplify the notation $\Delta_{n,t\tau}(s, x)$, $Z_{i,t\tau}$, and $R_{i,t\tau}(x)$ will be written as $\Delta_n(s, x)$, Z_i and $R_i(x)$, respectively.

We are going to obtain the rate of convergence of $\sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\Delta_n(s, x)|$ in probability, where the set $[-M_n, M_n]$ expands with n . The interest in $\Delta_n(s, x)$ comes from the representation (A.20) below. Bernstein's inequality is used to obtain the rate of convergence

of the supremum. Although the use of Bernstein's inequalities to obtain supremum rates of convergence of kernel estimators is standard (see for example Newey, 1994), the proof below is different from the usual ones in several aspects. As mentioned in the main text, the proof is complicated by the fact that the quantity of interest ($s^{-1}[\hat{\phi}_U^S(s|x) - \phi_U(s|x)]$ or $\Delta_n(s, x)$) exhibits different behavior for small, intermediate, and large values of $|s|$. Yet our interest is in obtaining a uniform rate over the values of $|s|$.

The values of s in the neighborhood of zero cause problems because of the s^{-1} term. This is the reason we consider sums of $s^{-1}(e^{isZ_i} - E[e^{isZ_i}|\mathcal{G}_x^{t\tau}])R_i(x)$ rather than of the more usual $s^{-1}e^{isZ_i}R_i(x)$. The variance of the former expression is shown to be bounded for all $s \neq 0$, while the variance of the latter is not bounded when s approaches zero.

When $|s|$ is large the problems are of a different sort. As mentioned in the main text, the bias of kernel estimators of such quantities as $E[\exp\{isY_{it}\}|X_{it} = x]$ and $E[\exp\{isZ_i\}|\mathcal{G}_x^{t\tau}]$ is growing with $|s|$.

Part 2 of the proof bounds $|\Delta_n(s, x) - E[\Delta_n(s, x)]|$ in probability. Part 3 provides the bound for the bias term $|E[\Delta_n(s, x)]|$. Part 4 combines these results to derive the result of the lemma.

2a. Consider the "stochastic" part, i.e. $\Delta_n(s, x) - E[\Delta_n(s, x)]$. Define $\rho_n = (\log(n)/(nh_U^{2p}))^{1/2}$ and $\eta(z, s, x) = s^{-1}(e^{isz} - E[e^{isZ_i}|\mathcal{G}_x^{t\tau}])$ for all $s \neq 0$ and by continuity take $\eta(z, 0, x) = \lim_{s \rightarrow 0} \eta(z, s, x) = i(z - E[Z_i|\mathcal{G}_x^{t\tau}])$. Note that

$$\eta(z, s, x) = i \int_0^s (\zeta/s) \left(z e^{i\zeta z} - E[Z_i e^{i\zeta Z_i}|\mathcal{G}_x^{t\tau}] \right) d\zeta \quad (\text{A.6})$$

and that $\Delta_n(s, x) = n^{-1} \sum_{i=1}^n \eta(\bar{Z}_{in}, s, x) R_i(x)$.

Define the random variable $\bar{Z}_{in} = Z_i 1\{|Z_i| \leq \log(n)n^{1/\gamma}\}$ and function

$$\bar{\Delta}_n(s, x) = n^{-1} \sum_{i=1}^n \eta(\bar{Z}_{in}, s, x) R_i(x).$$

By the triangle inequality for all s and x

$$\begin{aligned} |\Delta_n(s, x) - E[\Delta_n(s, x)]| &\leq |\Delta_n(s, x) - \bar{\Delta}_n(s, x)| \\ &+ |\bar{\Delta}_n(s, x) - E[\bar{\Delta}_n(s, x)]| + |E[\bar{\Delta}_n(s, x)] - E[\Delta_n(s, x)]|. \end{aligned} \quad (\text{A.7})$$

Part 2b of the proof shows that $\sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\bar{\Delta}_n(s, x) - E[\bar{\Delta}_n(s, x)]| = O_p(\rho_n)$ using Bernstein's inequality. Parts 2c and 2d show that $\sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |E[\Delta_n(s, x)] - E[\bar{\Delta}_n(s, x)]| = O(\rho_n)$ and $\sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\Delta_n(s, x) - \bar{\Delta}_n(s, x)| = O_p(\rho_n)$, respectively. Combined, these results prove that $\sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\Delta_n(s, x) - E[\Delta_n(s, x)]| = O_p(\rho_n)$.

2b. Note that for $s \in [-1, 1]$ we have $|\eta(z, s, x)| \leq |z| + E[|Z_i||\mathcal{G}_x^{t\tau}]$, where $E[|Z_i||\mathcal{G}_x^{t\tau}]$ is bounded by Assumption 3. For $s \in (-\infty, -1] \cup [1, \infty)$ we have $|\eta(\cdot)| \leq 2$ from the definition. Denote $C_\eta = \max\{2, \sup_{x \in \mathcal{X}} E[|Z_i||\mathcal{G}_x^{t\tau}]\}$ and notice that

$$\begin{aligned} E\left[|\eta(\bar{Z}_{in}, s, x) R_i(x)|^2\right] &\leq E\left[(|\bar{Z}_{in}| + C_\eta)^2 R_i^2(x)\right] \\ &\leq E\left[E\left[(|Z_i| + C_\eta)^2 \mid X_{it}, X_{i\tau}\right] R_i^2(x)\right] \leq CE[R_i^2(x)] \leq Ch_U^{-2p}, \end{aligned} \quad (\text{A.8})$$

where the first inequality follows from the bounds given under equation (A.6), the third inequality follows from Assumption 3 and the last inequality follows from Assumption 6.

It follows from (A.6) that

$$|\overline{\Delta}_n(s_1, x) - \overline{\Delta}_n(s_2, x)| \leq \left(\log(n) n^{1/\gamma} + C_\eta \right) h_U^{-2p} |s_1 - s_2|. \quad (\text{A.9})$$

Since $\sup_{(s,x) \in [-\varsigma, \varsigma] \times \mathcal{X}} |\eta(\overline{Z}_{in}, s, x)| = \varsigma (\log(n) n^{1/\gamma} + C)$ and $\sup_{(s,x) \in (-\infty, -\varsigma] \cup [\varsigma, \infty) \times \mathcal{X}} |\eta(\overline{Z}_{in}, s, x)| = 2\varsigma^{-1}$, one obtains

$$\sup_{(s,x) \in \mathbb{R} \times \mathcal{X}} |\eta(\overline{Z}_{in}, s, x)| \leq C \left(\log(n) n^{1/\gamma} + 1 \right). \quad (\text{A.10})$$

Next, for $s \in [-\varsigma, \varsigma]$ from (A.6)

$$\eta(z, s, x_1) - \eta(z, s, x_2) = \mathbf{i} \int_0^s (\zeta/s) \left(E \left[Z_i e^{\mathbf{i}\zeta Z_i} | \mathcal{G}_{x_1}^{t\tau} \right] - E \left[Z_i e^{\mathbf{i}\zeta Z_i} | \mathcal{G}_{x_2}^{t\tau} \right] \right) d\zeta.$$

Note that

$$\begin{aligned} \mathbf{i} E \left[Z_i e^{\mathbf{i}s Z_i} | \mathcal{G}_x^{t\tau} \right] &= \mathbf{i} E \left[(U_{it} - U_{i\tau}) e^{\mathbf{i}s(U_{it} - U_{i\tau})} | \mathcal{G}_x^{t\tau} \right] \\ &= \frac{\partial \phi_{U_{it}}(s|x)}{\partial s} \phi_{U_{i\tau}}(-s|x) - \phi_{U_{it}}(s|x) \frac{\partial \phi_{U_{i\tau}}(-s|x)}{\partial s}. \end{aligned}$$

Then, due to Assumption 4 function $E \left[Z_i e^{\mathbf{i}s Z_i} | \mathcal{G}_x^{t\tau} \right]$ belongs to $\mathcal{D}_{\mathcal{X}}^{[-\varsigma, \varsigma]}(\bar{k})$ and therefore $\sup_{z, s \in \mathbb{R} \times [-\varsigma, \varsigma]} |\eta(z, s, x_1) - \eta(z, s, x_2)| \leq \varsigma C \|x_1 - x_2\|^{\bar{k}}$. When $s \in (-\infty, -\varsigma] \cup [\varsigma, \infty)$ one obtains

$$\begin{aligned} |\eta(z, s, x_1) - \eta(z, s, x_2)| &= |s|^{-1} |\phi_{U_{it}}(s|x_1) \phi_{U_{i\tau}}(-s|x_1) - \phi_{U_{it}}(s|x_2) \phi_{U_{i\tau}}(-s|x_2)| \\ &\leq C \|x_1 - x_2\|^{\bar{k}}, \end{aligned}$$

where the inequality follows from Assumption 4, boundedness of characteristic functions and the fact that $|s|^{-1} \leq \varsigma^{-1}$. Thus, there is a constant $C > 0$ such that for all $(x_1, x_2) \in \mathcal{X} \times \mathcal{X}$

$$\sup_{z, s \in \mathbb{R} \times \mathbb{R}} |\eta(z, s, x_1) - \eta(z, s, x_2)| \leq C \|x_1 - x_2\|^{\bar{k}}. \quad (\text{A.11})$$

Now for all s :

$$\begin{aligned} &|\overline{\Delta}_n(s, x_1) - \overline{\Delta}_n(s, x_2)| \\ &\leq n^{-1} \sum_{i=1}^n [|\eta(\overline{Z}_{in}, s, x_1) - \eta(\overline{Z}_{in}, s, x_2)| |R_i(x_1)| + |\eta(\overline{Z}_{in}, s, x_2)| |R_i(x_1) - R_i(x_2)|] \\ &\leq C \left(h_U^{-2p} \|x_1 - x_2\|^{\bar{k}} + \log(n) n^{1/\gamma} h_U^{-2p-1} \|x_1 - x_2\| \right), \end{aligned} \quad (\text{A.12})$$

where the first inequality follows from "add and subtract" and the triangle inequality and the second inequality follows from (A.10), (A.11), and that for each n function $R_i(x)$ is bounded and Lipschitz by Assumption 6.

Take a sequence $N_n = n^{C_N}$ for some $C_N > 0$. Consider the grid $\{s_r\}_{r=-N_n+1}^{N_n}$ with $s_r = ((r - 1/2)/N_n) M_n$. Since $\mathcal{X} \subset \mathbb{R}^p$ is bounded, there is a positive constant $C_{\mathcal{X}}$ and a sequence of grids of $L_n = n^{C_L}$ points $\{x_q\}_{q=1}^{L_n} \subset \mathcal{X}$, such that $\sup_{x \in \mathcal{X}} \min_{q \in \{1, \dots, L_n\}} \|x - x_q\| < C_{\mathcal{X}} L_n^{-1/p}$. For any $s \in [-M_n, M_n]$ define $r^*(s)$ to be any element of $\{-N_n + 1, \dots, N_n\}$ such that $|s_{r^*(s)} - s| \leq M_n/N_n$. Define $q^*(x)$ similarly. Then for large n ,

$$\begin{aligned}
& \sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\overline{\Delta}_n(s, x) - E[\overline{\Delta}_n(s, x)]| \\
\leq & \sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\overline{\Delta}_n(s, x) - E[\overline{\Delta}_n(s, x)] - (\overline{\Delta}_n(s_{r^*(s)}, x) - E[\overline{\Delta}_n(s_{r^*(s)}, x)])| \\
& + \sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\overline{\Delta}_n(s_{r^*(s)}, x) - E[\overline{\Delta}_n(s_{r^*(s)}, x)] \\
& \quad - (\overline{\Delta}_n(s_{r^*(s)}, x_{q^*(x)}) - E[\overline{\Delta}_n(s_{r^*(s)}, x_{q^*(x)})])| \\
& + \max_{-N_n+1 \leq r \leq N_n, 1 \leq q \leq L_n} |\overline{\Delta}_n(s_r, x_q) - E[\overline{\Delta}_n(s_r, x_q)]| \\
\leq & C \log(n) n^{1/\gamma} h_U^{-2p} M_n/n^{C_N} + C \log(n) n^{1/\gamma} h_U^{-2p-1} C_{\mathcal{X}}/n^{C_L/p} \\
& + \max_{-N_n+1 \leq r \leq N_n, 1 \leq q \leq L_n} |\overline{\Delta}_n(s_r, x_q) - E[\overline{\Delta}_n(s_r, x_q)]|, \tag{A.13}
\end{aligned}$$

where the first inequality follows by the triangle inequality, and the second inequality follows from equations (A.9) and (A.12) and that $|E[Z_1] - E[Z_2]| \leq E[|Z_1 - Z_2|]$ for any random variables Z_1 and Z_2 .

Take the constants C_N and C_L sufficiently large so that the first two terms in the last line of (A.13) converge to zero at a rate, faster than ρ_n . Then, for large enough $a > 0$

$$\begin{aligned}
& \Pr \left\{ \sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\overline{\Delta}_n(s, x) - E[\overline{\Delta}_n(s, x)]| > (\sqrt{2} + 1) a \rho_n \right\} \\
\leq & \Pr \left\{ \max_{-N_n+1 \leq r \leq N_n, 1 \leq q \leq L_n} |\overline{\Delta}_n(s_r, x_q) - E[\overline{\Delta}_n(s_r, x_q)]| > \sqrt{2} a \rho_n \right\} \\
= & \Pr \left\{ \max_{\substack{-N_n+1 \leq r \leq N_n, \\ 1 \leq q \leq L_n}} \left| \frac{1}{n} \sum_{i=1}^n (\eta(\overline{Z}_{in}, s_r, x_q) R_i(x) - E[\eta(\overline{Z}_{in}, s_r, x_q) R_i(x)]) \right| > \sqrt{2} a \rho_n \right\} \\
\leq & 2 \cdot 2N_n \cdot 2L_n \exp \left\{ - \frac{a^2 n^2 \rho_n^2 / 2}{n \sup_{(s,x) \in \mathbb{R} \times \mathcal{X}} V[\eta(\overline{Z}_{in}, s, x) R_i(x)] + a \log(n) n^{1/\gamma+1} h_U^{-2p} \rho_n / 3} \right\} \\
\leq & 8N_n L_n \exp \left\{ - \frac{a^2 \log(n) h_U^{-2p} / 2}{C h_U^{-2p} + a [\log(n)]^{1/2} n^{1/\gamma-1/2} h_U^{-3p} / 3} \right\} \leq 8n^{C_L+C_N-Ca^2} \rightarrow 0 \tag{A.14}
\end{aligned}$$

where the second inequality follows from Bernstein inequality applied to the real and imaginary parts separately, the third inequality follows from (A.8) and the definition of ρ_n , the fourth inequality follows from Lemma's condition that $[\log(n)]^{1/2} n^{1/\gamma-1/2} h_U^{-p} = o(1)$ (cf. Assumption 8(ii)), and the convergence follows by taking a large enough.

2c. It follows from (A.6) that

$$\begin{aligned}
& |E[\Delta_n(s, x)] - E[\bar{\Delta}_n(s, x)]| \\
& \leq |s| E[|\bar{Z}_{in} - Z_i| |R_i(x)|] \\
& = |s| E\left[1 \left\{ |Z_i| > \log(n) n^{1/\gamma} \right\} |Z_i| |R_i(x)|\right] \\
& \leq |s| E\left[1 \left\{ |Z_i|^{\gamma-1} > (\log(n))^{\gamma-1} n^{1-1/\gamma} \right\} (\log(n))^{\gamma-1} n^{1/\gamma-1} |Z_i|^\gamma |R_i(x)|\right] \\
& \leq C |s| (\log(n))^{\gamma-1} n^{1/\gamma-1} E[E[|Z_i|^\gamma |X_{it}, X_{i\tau}] |R_i(x)]] \\
& \leq CM_n (\log(n))^{\gamma-1} n^{1/\gamma-1} = O(\rho_n), \tag{A.15}
\end{aligned}$$

where the last inequality follows from Assumptions 3 and 6, and the equality follows from Lemma's conditions (cf. Assumption 8(iii)).

2d. By Markov inequality

$$\begin{aligned}
& \Pr \left\{ \sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\Delta_n(s, x) - \bar{\Delta}_n(s, x)| > a\rho_n \right\} \\
& \leq n \Pr \left\{ |Z_i| > \log(n) n^{1/\gamma} \right\} \leq C (\log(n))^{-\gamma}.
\end{aligned}$$

2e. Using equation (A.7) and the results of parts 2b, 2c, and 2d we conclude that

$$\sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\Delta_n(s, x) - E[\Delta_n(s, x)]| = O_p(\rho_n). \tag{A.16}$$

3. Now consider the "deterministic" part, i.e. $E[\Delta_n(s, x)]$. From (A.6):

$$E[\Delta_n(s, x)] = i \int_0^s \frac{\zeta}{s} E \left[\left(Z_i e^{i\zeta Z_i} + \psi_t^U(\zeta|x) \phi_{U_{i\tau}}(-\zeta|x) - \psi_\tau^U(-\zeta|x) \phi_{U_{it}}(\zeta|x) \right) R_i(x) \right] d\zeta,$$

where $\psi_t^U(s|x) = E[U_{it} e^{isU_{it}} | X_{it} = x]$ and the order of integration can be changed by Assumption 3 and Fubini's theorem. It follows from Assumption 4 that function $E[Z_i e^{isZ_i} | \mathcal{G}_x^{t\tau}]$ of x belongs to $\mathcal{D}_{\mathcal{X}}^{[-\varsigma, \varsigma]}(\bar{k})$. Note that

$$E \left[Z_i e^{isZ_i} R_i(x) \right] = E \left[\psi_{t\tau}^Y(s|X_{it}, X_{i\tau}) h_U^{-2p} K \left(\frac{X_{it}-x}{h_U} \right) K \left(\frac{X_{i\tau}-x}{h_U} \right) \right],$$

where

$$\begin{aligned}
& \psi_{t\tau}^Y(s|x_t, x_\tau) \\
& = E \left[(Y_{it} - Y_{i\tau}) e^{is(Y_{it}-Y_{i\tau})} \middle| (X_{it}, X_{i\tau}) = (x_t, x_\tau) \right] \\
& = E \left[(m(x_t, \alpha_i) - m(x_\tau, \alpha_i) + U_{it} - U_{i\tau}) e^{is(m(x_t, \alpha_i) - m(x_\tau, \alpha_i) + U_{it} - U_{i\tau})} \middle| x_t, x_\tau \right] \\
& = \psi_{t\tau}^m(s|x_t, x_\tau) \phi_{U_{it}}(s|x_t) \phi_{U_{i\tau}}(-s|x_\tau) \\
& \quad + \phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_t, x_\tau) \left(\psi_t^U(s|x_t) \phi_{U_{i\tau}}(-s|x_\tau) - \phi_{U_{it}}(s|x_t) \psi_\tau^U(-s|x_\tau) \right),
\end{aligned}$$

where

$$\psi_{t\tau}^m(s|x_t, x_\tau) = E \left[(m(x_t, \alpha_i) - m(x_\tau, \alpha_i)) e^{is(m(x_t, \alpha_i) - m(x_\tau, \alpha_i))} \middle| x_t, x_\tau \right].$$

Note that $\psi_{t\tau}^m(s|x, x) = 0$ for all $x \in \mathcal{X}$ and hence

$$E[\Delta_n(s, x)] = -i \int_0^s (\zeta/s) (\varphi_{t\tau}^Y(x) - \psi_{t\tau}^Y(\zeta|x, x)) d\zeta,$$

where

$$\varphi_{t\tau}^Y(x) = \int \int \psi_{t\tau}^Y(s|\xi_1, \xi_2) f_{(X_{it}, X_{i\tau})}(\xi_1, \xi_2) h_U^{-2p} K\left(\frac{\xi_1 - x}{h_U}\right) K_{h_U}\left(\frac{\xi_1 - x}{h_U}\right) d\xi_1 d\xi_2. \quad (\text{A.17})$$

Note that

$$\begin{aligned} \psi_{t\tau}^m(s|x_t, x_\tau) &= -i \partial \phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_t, x_\tau) / \partial s, \\ \psi_t^U(s|x_t) &= -i \partial \phi_{U_{it}}(s|x_t) / \partial s. \end{aligned}$$

Then, $\psi_{t\tau}^Y(s|x_1, x_2) \in \mathcal{D}_{\mathcal{X} \times \mathcal{X}}^{[-\varsigma, \varsigma]}(\bar{k})$ for some constant C due to Assumptions 4 and 5 and the fact that $\psi_{t\tau}^m(s|x_t, x_\tau)$, $\psi_t^U(s|x_t)$, $\phi_{U_{it}}(s|x_t)$, and $\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_t, x_\tau)$ are bounded due to Assumption 3. Thus, using Assumptions 2 and 6 and equation (A.17) the standard argument yields

$$\begin{aligned} \sup_{(s,x) \in [-\varsigma, \varsigma] \times \mathcal{X}} |\varphi_{t\tau}^Y(x) - \psi_{t\tau}^Y(s|x, x)| &= O\left(h_U^{\bar{k}}\right), \text{ and hence} \\ \sup_{(s,x) \in [-\varsigma, \varsigma] \times \mathcal{X}} |E[\Delta_n(s, x)]| &= O\left(h_U^{-\bar{k}}\right). \end{aligned} \quad (\text{A.18})$$

Now consider any $s \in [-M_n, -\varsigma] \cup [\varsigma, M_n]$:

$$\begin{aligned} &E \left[\left(e^{isZ_i} - E \left[e^{isZ_i} | \mathcal{G}_x^{t\tau} \right] \right) R_i(x) \right] \\ &= \int \int \left[\phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x + z_1 h_U, x + z_2 h_U) \phi_{U_{it}}(s|x + z_1 h_U) \phi_{U_{i\tau}}(-s|x + z_2 h_U) \right. \\ &\quad \left. - \phi_{U_{it}}(s|x) \phi_{U_{i\tau}}(-s|x) \right] f_{X_{it}, X_{i\tau}}(x + z_1 h_U, x + z_2 h_U) K(z_1) K(z_2) dz_1 dz_2 \\ &\leq C \left(1 + |s|^{\bar{k}} \right) h_U^{\bar{k}} \end{aligned}$$

by Assumptions 4, 5, and 6. Therefore,

$$\sup_{(s,x) \in [-\bar{s}, \bar{s}] \times \mathcal{X}} |E[\Delta_n(s, x)]| = O\left(\left(1 + |\bar{s}|^{\bar{k}-1}\right) h_U^{\bar{k}}\right). \quad (\text{A.19})$$

4. Combining the results of steps 2 and 3 one obtains

$$\sup_{(s,x) \in [-M_n, M_n] \times \mathcal{X}} |\Delta_n(s, x)| = O_p\left([\log(n)/(nh_U^{2p})]^{1/2} + \left(1 + M_n^{\bar{k}-1}\right) h_U^{\bar{k}}\right).$$

Denote $\hat{f}_{(X_{it}, X_{i\tau})}(x, x) = n^{-1} \sum_{i=1}^n R_{i,t\tau}(x, h_U)$. Applying the result of Stone (1982) for

all pairs $t, \tau \neq t$,

$$\sup_{x \in \bar{\mathcal{X}}} \left| \hat{f}_{(X_{it}, X_{i\tau})}(x, x) - f_{(X_{it}, X_{i\tau})}(x, x) \right| = O_p \left([\log(n) / (nh_U^{2p})]^{1/2} + h_U^{\bar{k}} \right).$$

Note that

$$s^{-1} \left(\frac{\sum_{t < \tau} n^{-1} \sum_{i=1}^n e^{is(Y_{it} - Y_{i\tau})} R_{i,t\tau}(x)}{\sum_{t < \tau} n^{-1} \sum_{i=1}^n R_{i,t\tau}(x)} - \phi_U^2(s|x) \right) = \frac{\sum_{t < \tau} \Delta_{n,t\tau}(s, x)}{\sum_{t < \tau} n^{-1} \sum_{i=1}^n R_{i,t\tau}(x)}, \quad (\text{A.20})$$

and hence

$$\sup_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \left| s^{-1} \left(\left| \hat{\phi}_U^S(s|x) \right|^2 - \phi_U(s|x)^2 \right) \right| = O_p \left([\log(n) / (nh_U^{2p})]^{1/2} + \left(1 + M_n^{\bar{k}-1}\right) h_U^{\bar{k}} \right).$$

The conclusion of the lemma now follows from the fact that $\left| \hat{\phi}_U^S(s|x) \right| + \phi_U(s|x) \geq \phi_U(s|x)$ for all $(s, x) \in [-M_n, M_n] \times \bar{\mathcal{X}}$.

Lemma 5. *Suppose Assumptions ID(i)-(vi), and 1-7 hold. Assume that $nh_U \rightarrow \infty$, $h_U \rightarrow 0$, $[\log(n)]^{1/2} n^{1/\gamma-1/2} h_Y^{-d/2}(d)$, $M_n \rightarrow \infty$, and $(\log(n))^{\gamma-3/2} M_n h_Y^{d/2}(d) n^{1/\gamma-1/2} \rightarrow 0$ for $d \in \{p, 2p\}$. Then,*

$$\begin{aligned} \sup_{(s, x_t) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \left| s^{-1} \left(\hat{\phi}_{Y_{it}}(s|x_t) - \phi_{Y_{it}}(s|x_t) \right) \right| &= O_p(\rho_n^Y(p)) \text{ and} \\ \sup_{(s, x_{t\tau}) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \left| s^{-1} \left(\hat{\phi}_{Y_{it}}(s|x_{t\tau}) - \phi_{Y_{it}}(s|x_{t\tau}) \right) \right| &= O_p(\rho_n^Y(2p)), \end{aligned}$$

where $\rho_n^Y(d) = [\log(n) / (nh_Y^d(d))]^{1/2} + \left(1 + M_n^{\bar{k}-1}\right) h_Y^{\bar{k}}(d)$.

Proof. Analogous to the proof of Lemma 4. ■

Lemma 6. *Suppose Assumptions ASYM, ID(i)-(vi), and 1-7 hold. Assume that (i) $nh_U \rightarrow \infty$, $\max\{h_U, h_Y(p)\} \rightarrow 0$, (ii) $[\log(n)]^{1/2} n^{1/\gamma-1/2} \max\{h_U^{-p}, h_Y^{-p/2}(p)\} \rightarrow 0$, (iii) $M_n \rightarrow \infty$, $(\log(n))^{\gamma-3/2} M_n \max\{h_U^{-p}, h_Y^{-p/2}(p)\} n^{1/\gamma-1/2} \rightarrow 0$, and Assumption 8(v) holds. Then,*

$$\begin{aligned} &\sup_{(s, x_t) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \left| s^{-1} \left(\hat{\phi}_{U_{it}}(s|x) - \phi_{U_{it}}(s|x) \right) / \phi_{U_{it}}(s|x) \right| \\ &= O_p \left(\frac{\left(1 + M_n^{b_\phi}\right) \left[[\log(n) / (nh_U^{2p})]^{1/2} + \left(1 + M_n^{\bar{k}}\right) h_U^{\bar{k}} \right]}{\inf_{(s,x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \phi_{U_{it}}(s|x) \phi_{U_{i\tau}}(-s|x)} \right). \end{aligned}$$

Proof. Fix any $\tau \neq t$ and define the event $\mathcal{G}_x^{t\tau} = \{X_{it} = X_{i\tau} = x\}$. Define $R_{i,t\tau}(x) = h_U^{-2p} \kappa_i^{t,\tau}(x, h_U)$,

$$A_n^{t\tau}(s, x) = \frac{A_{n1}^{t\tau}(s, x)}{A_{n2}^{t\tau}(s, x)} = \frac{n^{-1} \sum_{i=1}^n Y_{it} e^{is(Y_{it} - Y_{i\tau})} R_{i,t\tau}(x)}{n^{-1} \sum_{i=1}^n e^{is(Y_{it} - Y_{i\tau})} R_{i,t\tau}(x)}, \quad B_n^{t\tau}(x) = \frac{\sum_{i=1}^n Y_{it} K_{h_Y}(X_{it} - x)}{\sum_{i=1}^n K_{h_Y}(X_{it} - x)},$$

$$A^{t\tau}(s, x) = \frac{A_1^{t\tau}(s, x)}{A_2^{t\tau}(s, x)} = \frac{E[Y_{it}e^{is(Y_{it}-Y_{i\tau})}|\mathcal{G}_x^{t\tau}]f_{(X_{it}, X_{i\tau})}(x, x)}{E[e^{is(Y_{it}-Y_{i\tau})}|\mathcal{G}_x^{t\tau}]f_{(X_{it}, X_{i\tau})}(x, x)}, \text{ and } B^{t\tau}(x) = E[Y_{it}|\mathcal{G}_x^{t\tau}].$$

Similar to step 2 of the proof of Lemma 4 one can show that

$$\begin{aligned} & \sup_{(s, x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} \left| n^{-1} \sum_{i=1}^n \left(Y_{it} e^{is(Y_{it}-Y_{i\tau})} R_{i,t\tau}(x) - E \left[Y_{it} e^{is(Y_{it}-Y_{i\tau})} R_{i,t\tau}(x) \right] \right) \right| \\ &= O_p \left(\left[\log(n) / (nh_U^{2p}) \right]^{1/2} \right). \end{aligned} \quad (\text{A.21})$$

Note that

$$\begin{aligned} & E \left[Y_{it} e^{is(Y_{it}-Y_{i\tau})} | x_{t\tau} \right] \\ &= E \left[Y_{it} e^{is(Y_{it}-Y_{i\tau})} | (X_{it}, X_{i\tau}) = (x_t, x_\tau) \right] \\ &= E \left[(m(x_t, \alpha_i) + U_{it}) e^{is(m(x_t, \alpha_i) - m(x_\tau, \alpha_i) + U_{it} - U_{i\tau})} | (X_{it}, X_{i\tau}) = x_t, x_\tau \right] \\ &= E \left[m(x_t, \alpha_i) e^{is(m(x_t, \alpha_i) - m(x_\tau, \alpha_i))} | x_t, x_\tau \right] \phi_{U_{it}}(s|x_\tau) \phi_{U_{i\tau}}(-s|x_\tau) \\ &\quad + \phi_{m(x_t, \alpha_i) - m(x_\tau, \alpha_i)}(s|x_t, x_\tau) E \left[U_{it} e^{isU_{it}} | X_{it} = x_t \right] \phi_{U_{i\tau}}(-s|x_\tau). \end{aligned}$$

Then, function $E[Y_{it}e^{is(Y_{it}-Y_{i\tau})}|x_{t\tau}]f_{(X_{it}, X_{i\tau})}(x_{t\tau})$ of $x_{t\tau}$ belongs to $\mathcal{D}_{\bar{\mathcal{X}} \times \mathcal{X}}^{\mathbb{R}}(\bar{k})$ due to Assumptions 2-5. Thus, using the standard argument,

$$\begin{aligned} & \sup_{(s, x) \in [-\bar{s}, \bar{s}] \times \bar{\mathcal{X}}} \left| E \left[Y_{it} e^{is(Y_{it}-Y_{i\tau})} R_{i,t\tau}(x, h_U) \right] - E \left[Y_{it} e^{is(Y_{it}-Y_{i\tau})} | \mathcal{G}_x^{t\tau} \right] f_{(X_{it}, X_{i\tau})}(x, x) \right| \\ &= O \left(\left(1 + \bar{s}^{\bar{k}} \right) h_U^{\bar{k}} \right). \end{aligned} \quad (\text{A.22})$$

Denote $\rho_{An} = [\log(n)/(nh_U^{2p})]^{1/2} + (1 + \bar{s}^{\bar{k}})h_U^{\bar{k}}$ and

$$\Delta A_{kn}^{t\tau}(s, x) = A_{kn}^{t\tau}(s, x) - A_k^{t\tau}(s, x), \quad k = 1, 2.$$

Combine (A.21) and (A.22) to obtain

$$\sup_{(s, x) \in [-\bar{s}, \bar{s}] \times \bar{\mathcal{X}}} |\Delta A_{1n}^{t\tau}| = O_p(\rho_{An}).$$

Similarly, one can show that

$$\sup_{(s, x) \in [-\bar{s}, \bar{s}] \times \bar{\mathcal{X}}} |\Delta A_{2n}^{t\tau}(s, x)| = O_p(\rho_{An}).$$

Define $\Delta B_n^{t\tau}(x) = B_n^{t\tau}(x) - B^{t\tau}(x)$. Then, it follows from Stone (1982) that

$$\sup_{x \in \bar{\mathcal{X}}} |\Delta B_n^{t\tau}(x)| = O_p \left(\left[\log(n) / (nh_Y^p(p)) \right]^{1/2} + h_Y^{\bar{k}}(p) \right),$$

Then, $\sup_{x \in \bar{\mathcal{X}}} |\Delta B_n^{t\tau}(x)| = O_p(\rho_{An})$ holds by Assumption 8(v).

From the identity $\hat{A}_1/\hat{A}_2 - A_1/A_2 = ((\hat{A}_1 - A_1)A_2 - A_1(\hat{A}_2 - A_2))/(A_2\hat{A}_2)$ one obtains

$$\frac{A_{n1}^{t\tau}(s, x)}{A_{n2}^{t\tau}(s, x)} - \frac{A_1^{t\tau}(s, x)}{A_2^{t\tau}(s, x)} = \left(\frac{\Delta A_{n1}^{t\tau}(s, x)}{A_2^{t\tau}(s, x)} + \frac{A_1^{t\tau}(s, x) \Delta A_{n2}^{t\tau}(s, x)}{A_2^{t\tau}(s, x)^2} \right) \varrho_n(s, x), \quad (\text{A.23})$$

where $\varrho_n(s, x) = A_2^{t\tau}(s, x)/A_{n2}^{t\tau}(s, x) = (1 + \Delta A_{n2}^{t\tau}(s, x)/A_2^{t\tau}(s, x))^{-1}$. Note that

$$\begin{aligned} A_1^{t\tau}(s, x) &= E \left[(m(x, \alpha_i) + U_{it}) e^{is(U_{it} - U_{i\tau})} | \mathcal{G}_x^{t\tau} \right] \\ &= \left(E [m(x, \alpha_i) | \mathcal{G}_x^{t\tau}] \phi_{U_{it}}(s|x) + E [U_{it} e^{isU_{it}} | X_{it} = x] \right) \phi_{U_{i\tau}}(-s|x), \\ A_2^{t\tau}(s, x) &= \phi_{U_{it}}(s|x) \phi_{U_{i\tau}}(-s|x). \end{aligned}$$

Thus

$$\frac{A_1^{t\tau}(s, x)}{A_2^{t\tau}(s, x)} = E [m(x, \alpha_i) | \mathcal{G}_x^{t\tau}] - i \frac{\partial \ln \phi_{U_{it}}(s|x)}{\partial s}.$$

where $\ln(\cdot)$ is the principal value of complex logarithm. Note that $\varrho_n(s, x) = 1 + o_p(1)$ when the parenthesis term in (A.23) is $o_p(1)$. Denote

$$\Delta_n^{AB, t\tau}(s, x) = [A_n^{t\tau}(s, x) + B_n^{t\tau}(x)] - [A^{t\tau}(s, x) + B^{t\tau}(x)],$$

Then, equation (A.23) and Assumption ASYM yield

$$|\Delta_n^{AB, t\tau}(s, x)| \leq C \left(1 + |s|^{b_\phi} \right) \frac{|\Delta A_{n1}^{t\tau}(s, x)| + |\Delta A_{n2}^{t\tau}(s, x)|}{|A_2^{t\tau}(s, x)|} \varrho_n(s, x) + |\Delta B_n^{t\tau}(x)|,$$

hence

$$\inf_{(s, x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\Delta_n^{AB, t\tau}(s, x)| = O_p \left(\frac{(1 + M_n^{b_\phi}) [\log(n)/(nh_U^{2p})]^{1/2} + (1 + M_n^{\bar{k}}) \bar{h}_U^{\bar{k}}}{\inf_{(s, x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\phi_{U_{it}}(s|x) \phi_{U_{i\tau}}(-s|x)|} \right). \quad (\text{A.24})$$

Note that

$$\hat{\phi}_{U_{it}}(s|x) = \phi_{U_{it}}(s|x) \frac{1}{T-1} \sum_{\tau=1, \tau \neq t}^T \exp \left(i \int_0^s \Delta_n^{AB, t\tau}(\zeta, x) d\zeta \right).$$

Hence, for all $s \in [M_n, 0) \cup (0, M_n]$,

$$\begin{aligned} & \left| s^{-1} \left(\hat{\phi}_{U_{it}}(s|x) - \phi_{U_{it}}(s|x) \right) / \phi_{U_{it}}(s|x) \right| \\ & \leq \frac{1}{T-1} \sum_{\tau=1, \tau \neq t}^T |s|^{-1} \left| \exp \left(i \int_0^s \Delta_n^{AB, t\tau}(\zeta, x) d\zeta \right) - 1 \right| \\ & \leq C \frac{1}{T-1} \sum_{\tau=1, \tau \neq t}^T |s|^{-1} \int_0^s |\Delta_n^{AB, t\tau}(\zeta, x)| d\zeta \\ & \leq C \frac{1}{T-1} \sum_{\tau=1, \tau \neq t}^T \inf_{(s, x) \in [-M_n, M_n] \times \bar{\mathcal{X}}} |\Delta_n^{AB, t\tau}(s, x)|. \end{aligned}$$

Thus the conclusion of the lemma follows from (A.24). \blacksquare

Proof of Theorem 3. Note that

$$\begin{aligned}\hat{F}_m(\omega|x_t) - F_m(\omega|x_t) &= \frac{1}{2} + \int_{-1/h_w}^{1/h_w} \frac{e^{-is\omega}}{2\pi is} \phi_w(h_w s) \frac{\phi_{Y_{it}}(s|x_t)}{\phi_{U_{it}}(s|x_t)} ds - F_m(\omega|x_t) \\ &\quad + \int_{-1/h_w}^{1/h_w} \frac{e^{-is\omega}}{2\pi is} \phi_w(h_w s) \left[\frac{\hat{\phi}_{Y_{it}}(s|x_t)}{\hat{\phi}_{U_{it}}(s|x_t)} - \frac{\phi_{Y_{it}}(s|x_t)}{\phi_{U_{it}}(s|x_t)} \right] ds \\ &= A_{1n} + A_{2n},\end{aligned}$$

where $\hat{\phi}_{U_{it}}(s|x_t)$ denotes $\hat{\phi}_{U_{it}}^S(s|x_t)$ or $\hat{\phi}_{U_{it}}^{AS}(s|x_t)$ when, respectively, SYM or ASYM holds. The first term satisfies

$$\begin{aligned}A_{1n} &= \frac{1}{2} + \int_{-1/h_w}^{1/h_w} \frac{e^{-is\omega}}{2\pi is} \phi_w(h_w s) \phi_m(s|x_t) ds - F_m(\omega|x_t) \\ &= \int_{-\infty}^{\infty} F_m(\omega - u|x_t) K_w(h_w u) du - F_m(\omega|x_t).\end{aligned}$$

Thus $\sup_{(\omega,x) \in \mathbb{R} \times \mathcal{X}} |A_{1n}| = O(h_w^{k_F})$ due to Assumptions 1 and 7 by the standard argument. Then, similar to the expansion in (A.23) (and also similar to equation (A.33) in Schennach (2004b)) we obtain

$$\begin{aligned}\frac{1}{s} \left[\frac{\hat{\phi}_{Y_{it}}(s|x_t)}{\hat{\phi}_{U_{it}}(s|x_t)} - \frac{\phi_{Y_{it}}(s|x_t)}{\phi_{U_{it}}(s|x_t)} \right] &= \\ \left[\frac{\hat{\phi}_{Y_{it}}(s|x_t) - \phi_{Y_{it}}(s|x_t)}{s\phi_{U_{it}}(s|x_t)} - \frac{\phi_{Y_{it}}(s|x_t) [\hat{\phi}_{U_{it}}(s|x_t) - \phi_{U_{it}}(s|x_t)]}{s\phi_{U_{it}}(s|x_t)^2} \right] &\left(1 + \frac{\hat{\phi}_{U_{it}}(s|x_t) - \phi_{U_{it}}(s|x_t)}{\phi_{U_{it}}(s|x_t)} \right)^{-1}.\end{aligned}$$

Note that $|\phi_{Y_{it}}(s|x_t)/\phi_{U_{it}}(s|x_t)| = |\phi_m(s|x_t)|$ and hence $|\phi_m(s|x_t)| < C$ for $|s| \leq 1$ and $|\phi_m(s|x_t)| < C |s|^{-(k_F-1)}$ for $|s| > 1$ due to Assumption 1.

Denote

$$\varrho_n = \sup_{(s,x_t) \in [-h_w^{-1}, h_w^{-1}] \times \bar{\mathcal{X}}} \left| \frac{\hat{\phi}_{U_{it}}(s|x_t) - \phi_{U_{it}}(s|x_t)}{\phi_{U_{it}}(s|x_t)} \right|.$$

When Assumption ASYM holds, Lemma 6 gives

$$\varrho_n = O_p \left(h_w^{-b_\phi - 1} \left[[\log(n)/(nh_U^{2p})]^{1/2} + h_w^{-\bar{r}} h_U^{\bar{k}} \right] \chi^2(h_w^{-1}) \right),$$

with $\bar{r} = \bar{k}$. If Assumption SYM holds Lemma 4 yields the same result with $b_\phi = 0$ and $\bar{r} = \bar{k} - 1$. Assumption 8(iv) ensures that $\varrho_n = o_p(1)$.

Then,

$$\begin{aligned}
& \sup_{(\omega, x_t) \in \mathbb{R} \times \bar{\mathcal{X}}} \left| \int_{-h_w^{-1}}^{h_w^{-1}} \frac{e^{-isw}}{2\pi is} \phi_w(h_ws) \left[\frac{\hat{\phi}_{Y_{it}}(s|x_t)}{\hat{\phi}_{U_{it}}(s|x_t)} - \frac{\phi_{Y_{it}}(s|x_t)}{\phi_{U_{it}}(s|x_t)} \right] ds \right| \\
& \leq (C + o_p(1)) \left(\sup_{(s, x_t) \in [-h_w^{-1}, h_w^{-1}] \times \bar{\mathcal{X}}} s^{-1} \left| \hat{\phi}_{Y_{it}}(s|x_t) - \phi_{Y_{it}}(s|x_t) \right| \right) h_w^{-1} \chi(h_w^{-1}) \\
& \quad + (C + o_p(1)) \left(\sup_{(s, x_t) \in [-h_w^{-1}, h_w^{-1}] \times \bar{\mathcal{X}}} |\phi_m(s|x_t)| \left| \frac{\hat{\phi}_{U_{it}}(s|x_t) - \phi_{U_{it}}(s|x_t)}{s\phi_{U_{it}}(s|x_t)} \right| \right) h_w^{-1} \\
& = (C + o_p(1)) (B_{1n} + B_{2n}),
\end{aligned}$$

By Lemma 5 one obtains

$$B_{1n} = O_p \left(\left[[\log(n) / (nh_Y^p(p))]^{1/2} + h_w^{-(\bar{k}-1)} h_Y^{\bar{k}}(p) \right] h_w^{-1} \chi(h_w^{-1}) \right),$$

while Lemmas 4 and 6 give

$$\begin{aligned}
B_{2n} & \leq O_p \left(\left([\log(n) / (nh_U^{2p})]^{1/2} + h_U^{\bar{r}-1} \right) h_w^{-1} \right) \\
& \quad + \sup_{\bar{s} \in [1, h_w^{-1}]} \left[\bar{s}^{1-k_F} \sup_{(s, x_t) \in [-\bar{s}, \bar{s}] \times \bar{\mathcal{X}}} \left| \frac{\hat{\phi}_{U_{it}}(s|x_t) - \phi_{U_{it}}(s|x_t)}{s\phi_{U_{it}}(s|x_t)} \right| \right] h_w^{-1} \\
& = O_p \left(\left([\log(n) / (nh_U^{2p})]^{1/2} + h_U^{\bar{r}-1} \right) h_w^{-1} \right) \\
& \quad + O_p \left(\sup_{\bar{s} \in [1, h_w^{-1}]} \left[\bar{s}^{1-k_F} \chi^2(\bar{s}) \bar{s}^{b_\phi} \left([\log(n) / (nh_U^{2p})]^{1/2} + \bar{s}^{\bar{r}} h_U^{\bar{k}} \right) \right] h_w^{-1} \right) \\
& = O_p \left(\max \left\{ 1, h_w^{k_F-1-b_\phi} \chi^2(h_w^{-1}) \right\} \left([\log(n) / (nh_U^{2p})]^{1/2} + h_w^{-\bar{r}} h_U^{\bar{k}} \right) h_w^{-1} \right).
\end{aligned}$$

Therefore, $\sup_{(\omega, x_t) \in \mathbb{R} \times \bar{\mathcal{X}}} \left| \hat{F}_m(\omega|x_t) - F_m(\omega|x_t) \right| = O_p(\beta_n(p))$. In exactly the same way (only the term B_{1n} differs) one obtains that $\sup_{(\omega, x_{t\tau}) \in \mathbb{R} \times \bar{\mathcal{X}} \times \bar{\mathcal{X}}} \left| \hat{F}_m(\omega|x_{t\tau}) - F_m(\omega|x_{t\tau}) \right| = O_p(\beta_n(2p))$. ■

Proof of Theorem 4. 1. We can now write down optimal h_U and $h_Y(d)$ for $d \in \{p, 2p\}$ as functions of h_w . We get

$$h_Y^*(h_w, d) \sim \left[(\log(n)/n)^{1/2} h_w^{\bar{k}-1} \right]^{2/(2\bar{k}+d)}, \quad (\text{A.25})$$

$$h_U^*(h_w) \sim \left[(\log(n)/n)^{1/2} h_w^{\bar{r}} \right]^{1/(\bar{k}+p)}. \quad (\text{A.26})$$

With optimal h_Y and h_U one can simplify $\beta_n(d)$:

$$\begin{aligned} \beta_n(d) &\sim h_w^{k_F} + (\log(n)/n)^{\frac{\bar{k}}{2\bar{k}+d}} h_w^{-\frac{(\bar{k}-1)d}{2\bar{k}+d}-1} \chi(h_w^{-1}) \\ &\quad + (\log(n)/n)^{\frac{\bar{k}}{2(\bar{k}+p)}} h_w^{-\frac{\bar{r}p}{\bar{k}+p}-1} \max\{1, h_w^{k_F-1-b_\phi} \chi^2(h_w^{-1})\} = T_1 + T_2 + T_3. \end{aligned} \quad (\text{A.27})$$

Under Assumption OS we have $\chi(h_w^{-1}) \leq C h_w^{-\bar{\lambda}}$ and $b_\phi = 0$.

2. Consider the case $h_w^{k_F-1} \chi^2(h_w^{-1}) \rightarrow \infty$ as $h_w \rightarrow 0$, i.e. $k_F < 2\bar{\lambda} + 1$. The stochastic order of the terms T_1 and T_2 is the same when $h_w \sim (\ln(n)/n)^{\varkappa_{OS1}(d)}$, where

$$\varkappa_{OS1}(d) = \frac{\bar{k}/(2\bar{k}+d)}{k_F + \bar{\lambda} + 1 + (\bar{k}-1)d/(2\bar{k}+d)}.$$

Similarly, the stochastic order of the terms T_1 and T_3 is the same when $h_w \sim (\log(n)/n)^{\varkappa_{OS2}}$, where

$$\varkappa_{OS2} = \frac{\bar{k}/(2\bar{k}+2p)}{2(\bar{\lambda}+1) + \bar{r}p/(\bar{k}+p)},$$

which does not depend on d , since the terms T_1 and T_3 do not depend on d . When $k_F \geq \tilde{k}_F(d)$ holds, it is easy to show that $\beta_n(d)$ converges to zero at the fastest rate when the stochastic order of terms T_1 and T_2 is balanced (and hence $h_w \sim (\ln(n)/n)^{\varkappa_{OS1}(d)}$). In this case the stochastic order of the term T_3 is no bigger (smaller, if $k_F > \tilde{k}_F(d)$) than that of T_3 . Similarly, under the condition $k_F < \tilde{k}_F(d)$ one can show that $\beta_n(d)$ tends to zero at the fastest rate when the terms T_1 and T_3 are balanced (hence $h_w \sim (\log(n)/n)^{\varkappa_{OS2}}$). In this case the term T_2 is smaller (in the limit) than the terms T_1 and T_3 .

3. Now consider the case $k_F > 2\bar{\lambda} - 1$. When $d = 2p$ and $\bar{\lambda} \geq 1$ term T_2 always dominates term T_3 ; thus in this case the rate of convergence is $O_p\left((\ln(n)/n)^{k_F \varkappa_{OS1}(2p)}\right)$. When $d = p$, terms T_1 and T_3 have the same stochastic order when $h_w \sim (\log(n)/n)^{\varkappa_{OS3}}$

$$\varkappa_{OS3} = \frac{\bar{k}/(2\bar{k}+2p)}{k_F + 1 + \bar{r}p/(\bar{k}+p)},$$

while T_1 and T_2 are balanced when $h_w \sim (\log(n)/n)^{\varkappa_{OS1}(p)}$ and with these bandwidths term T_3 is larger than T_2 in the limit when $k_F > \bar{k}_F = \bar{\lambda}(2\bar{k}/p + 1) + \bar{k} - 2 - 2\bar{r}$. Notice also that if $\bar{\lambda}$ is close to zero term T_3 always dominates T_2 .

Define

$$\varkappa_{OS,U} = \frac{\bar{k}/(2\bar{k}+2p)}{\max\{k_F, 2\bar{\lambda} + 1\} + 1 + \bar{r}p/(\bar{k}+p)}.$$

We obtain that $\beta_n(d)$ converges to zero at the fastest rate when the bandwidth h_w satisfies

$h_w \sim (\ln(n)/n)^{\varkappa_{OS}(d, k_F)}$, where

$$\varkappa_{OS}(2p, k_F) = \begin{cases} \varkappa_{OS,U}, & \text{when } k_F < \tilde{k}_F(2p) \text{ and } k_F < 2\bar{\lambda} + 1, \\ \varkappa_{OS1}(2p), & \text{otherwise.} \end{cases}$$

$$\varkappa_{OS}(p, k_F) = \begin{cases} \varkappa_{OS,U}, & \text{when } k_F \leq \tilde{k}_F(p) \text{ and } k_F \leq 2\bar{\lambda} + 1, \\ \varkappa_{OS1}(p), & \text{when } k_F > \tilde{k}_F(p) \text{ and } k_F \leq 2\bar{\lambda} + 1, \\ \varkappa_{OS1}(p), & \text{when } k_F \leq \check{k}_F \text{ and } k_F > 2\bar{\lambda} + 1, \\ \varkappa_{OS,U}, & \text{when } k_F > \check{k}_F \text{ and } k_F > 2\bar{\lambda} + 1. \end{cases}$$

and hence $\beta_n(d) = O\left((\ln(n)/n)^{k_F \varkappa_{OS}(d, k_F)}\right)$ and the corresponding optimal bandwidths $h_Y(d)$ and h_U are given by (A.25) and (A.26), respectively. It is only left to check that Assumptions 8(ii)-(v) hold for optimal h_w , h_U , $h_Y(d)$ in steps 2 and 3 of the proof. A tedious but straightforward analysis shows that Assumption 9 is sufficient to ensure that the conditions of Assumption 8 are indeed satisfied for the optimal bandwidths. ■

Proof of Theorem 5. Under Assumption SS we have $\chi(h_w^{-1}) \leq Ch_w^{\lambda_0} \exp\left(h_w^{-\bar{\lambda}}/\underline{\mu}\right)$ and the rate (A.27) becomes

$$\begin{aligned} \beta_n(d) \sim & h_w^{k_F} + (\log(n)/n)^{\frac{\bar{k}}{2\bar{k}+d}} h_w^{-\frac{(\bar{k}-1)d}{2\bar{k}+d} - 1 - \bar{C}_2} \exp\left(h_w^{-\bar{\lambda}}/\bar{C}_3\right) \\ & + (\log(n)/n)^{\frac{\bar{k}}{2(\bar{k}+p)}} h_w^{-\frac{\bar{r}p}{\bar{k}+p} - 1} \max\left\{1, h_w^{k_F - 1 - b_\phi - 2\bar{C}_2} \exp\left(2h_w^{-\bar{\lambda}}/\bar{C}_3\right)\right\}. \end{aligned} \quad (\text{A.28})$$

It must be that $h_w \leq C(\ln(n))^{-1/\bar{\lambda}}$ for some positive constant C , otherwise $\exp\left(h_w^{-\bar{\lambda}}/\underline{\mu}\right)$ would grow with n at a faster than polynomial rate and $\hat{F}_m(\omega|x_{t\tau})$ would diverge. Thus, one immediately concludes that $\beta_n(d) = O\left((\ln n)^{-k_F/\bar{\lambda}}\right)$, for $d \in \{p, 2p\}$. One can take $h_w = C_{SS}(\ln n)^{-1/\bar{\lambda}}$, where the constant C_{SS} should be taken sufficiently small so that the term $h_w^{k_F}$ is not dominated by the other terms in equation (A.28) and Assumption 8(iv) holds. Assumptions 8(i),(iii), and (v) hold for the optimal $h_Y(d)$ and h_U when h_w is logarithmic. Finally, Assumption 10 ensures that Assumption 8(ii) holds. ■

Proof of Corollary 6. For example, see the proof of Theorem 3.1 in Ould-Saïd, Yahia, and Necir (2009). ■

Proof of Theorem 7. Define $\bar{\Omega}_{x\omega} = \bar{\mathcal{X}} \times [\delta, 1 - \delta]$ and take $\varepsilon_\omega = \min\{\delta/2, \varepsilon_{\mathcal{X}}/2\}$. Then the conditions of Corollary 6 are satisfied and the conclusion of the theorem follows. ■

Proof of Theorem 8. The assumptions of the theorem ensure that functions $Q_{\alpha_i|X_{it}}(q|x_t)$, $Q_{\alpha_i|X_{it}, X_{i\tau}}(q|x_t, x_\tau)$, and hence $\underline{\alpha}(q|x_t)$ and $\bar{\alpha}(q|x_t)$ are uniformly continuous in x_t and x_τ for all $(x_t, x_\tau) \in \mathcal{X} \times \mathcal{X}$, $q \in [\delta/2, 1 - \delta/2]$ and $t, \tau \neq t$. Thus there is an $\varepsilon_\omega > 0$ such that the conditions of Corollary 6 hold with the set $\bar{\Omega}_{x\omega} = S_{x,\alpha}(\delta)$. Then the conclusions of the theorem follow. ■

6.6 Results of Monte Carlo Study

		DESIGN I, $n = 2500, T = 2$										
h_Y	h_w		$\varepsilon_{it} \sim \text{Laplace}$					$\varepsilon_{it} \sim N(0, 1)$				
			0.875	0.75	0.5	0.25	0.125	0.875	0.75	0.5	0.25	0.125
0.2	0.2	<i>RIMSE</i>	1.06	0.65	0.38	0.42	1.02	0.73	0.49	0.36	0.42	0.80
		<i>RIBIAS</i> ²	0.73	0.35	0.14	0.19	0.76	0.44	0.21	0.07	0.16	0.58
		<i>RIVAR</i>	0.76	0.55	0.35	0.38	0.69	0.58	0.44	0.35	0.39	0.55
0.2	0.4	<i>RIMSE</i>	0.49	0.30	0.19	0.22	0.38	0.47	0.33	0.25	0.31	0.46
		<i>RIBIAS</i> ²	0.15	0.06	0.01	0.05	0.19	0.26	0.16	0.03	0.16	0.32
		<i>RIVAR</i>	0.47	0.29	0.19	0.22	0.32	0.39	0.29	0.25	0.27	0.34
0.4	0.2	<i>RIMSE</i>	0.99	0.63	0.31	0.40	1.03	0.79	0.52	0.35	0.45	1.03
		<i>RIBIAS</i> ²	0.72	0.41	0.14	0.26	0.80	0.52	0.27	0.13	0.21	0.76
		<i>RIVAR</i>	0.68	0.47	0.27	0.31	0.65	0.59	0.44	0.32	0.40	0.69
0.4	0.4	<i>RIMSE</i>	0.38	0.23	0.16	0.19	0.30	0.42	0.30	0.23	0.26	0.41
		<i>RIBIAS</i> ²	0.10	0.09	0.07	0.09	0.14	0.22	0.16	0.10	0.12	0.24
		<i>RIVAR</i>	0.36	0.22	0.14	0.17	0.26	0.36	0.25	0.21	0.23	0.33
0.4	0.6	<i>RIMSE</i>	0.38	0.27	0.13	0.27	0.39	0.46	0.34	0.17	0.32	0.46
		<i>RIBIAS</i> ²	0.33	0.23	0.07	0.23	0.35	0.39	0.29	0.07	0.25	0.39
		<i>RIVAR</i>	0.19	0.14	0.11	0.13	0.18	0.24	0.19	0.16	0.19	0.24
0.6	0.4	<i>RIMSE</i>	0.32	0.22	0.18	0.22	0.36	0.40	0.31	0.25	0.28	0.45
		<i>RIBIAS</i> ²	0.13	0.13	0.13	0.15	0.23	0.22	0.21	0.17	0.15	0.25
		<i>RIVAR</i>	0.30	0.18	0.13	0.15	0.27	0.34	0.24	0.19	0.23	0.37
0.6	0.6	<i>RIMSE</i>	0.38	0.29	0.16	0.27	0.42	0.45	0.36	0.19	0.29	0.45
		<i>RIBIAS</i> ²	0.35	0.26	0.12	0.25	0.39	0.40	0.32	0.13	0.24	0.38
		<i>RIVAR</i>	0.14	0.12	0.10	0.12	0.16	0.20	0.17	0.14	0.17	0.24
INFEASIBLE		<i>RIMSE</i>	0.21	0.17	0.15	0.17	0.21	0.21	0.17	0.15	0.16	0.21

Table 1. Design I, $n = 2500, T = 2$. Monte Carlo results of estimation of functions $\hat{h}(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. *RIMSE*, *RIBIAS*², and *RIVAR* stand for square roots of integrated Mean Squared Error, Squared Bias, and Variance, respectively.

DESIGN I, $n = 10000$, $T = 2$

h_Y	h_w		$\varepsilon_{it} \sim Laplace$					$\varepsilon_{it} \sim N(0, 1)$				
			0.875	0.75	0.5	0.25	0.125	0.875	0.75	0.5	0.25	0.125
0.2	0.2	<i>RIMSE</i>	0.84	0.49	0.21	0.32	0.78	0.62	0.39	0.24	0.32	0.70
		<i>RIBIAS</i> ²	0.67	0.33	0.08	0.22	0.62	0.41	0.20	0.05	0.15	0.54
		<i>RIVAR</i>	0.52	0.36	0.20	0.24	0.48	0.47	0.34	0.23	0.28	0.45
0.2	0.4	<i>RIMSE</i>	0.31	0.18	0.11	0.13	0.19	0.31	0.21	0.15	0.20	0.29
		<i>RIBIAS</i> ²	0.07	0.05	0.02	0.03	0.08	0.16	0.11	0.03	0.10	0.18
		<i>RIVAR</i>	0.30	0.17	0.10	0.12	0.18	0.27	0.19	0.15	0.17	0.23
0.4	0.2	<i>RIMSE</i>	0.85	0.49	0.19	0.33	0.82	0.68	0.43	0.24	0.36	0.88
		<i>RIBIAS</i> ²	0.73	0.37	0.12	0.28	0.72	0.49	0.27	0.13	0.24	0.70
		<i>RIVAR</i>	0.44	0.31	0.15	0.17	0.41	0.47	0.33	0.21	0.27	0.53
0.4	0.4	<i>RIMSE</i>	0.21	0.15	0.10	0.13	0.19	0.29	0.21	0.15	0.18	0.27
		<i>RIBIAS</i> ²	0.09	0.08	0.07	0.09	0.13	0.16	0.14	0.09	0.11	0.17
		<i>RIVAR</i>	0.19	0.12	0.07	0.09	0.14	0.25	0.17	0.12	0.15	0.21
0.4	0.6	<i>RIMSE</i>	0.34	0.24	0.09	0.24	0.37	0.38	0.28	0.11	0.26	0.40
		<i>RIBIAS</i> ²	0.32	0.23	0.07	0.23	0.36	0.35	0.26	0.07	0.23	0.37
		<i>RIVAR</i>	0.10	0.08	0.06	0.07	0.09	0.15	0.11	0.09	0.11	0.15
0.6	0.4	<i>RIMSE</i>	0.20	0.16	0.15	0.18	0.29	0.30	0.23	0.20	0.22	0.37
		<i>RIBIAS</i> ²	0.13	0.12	0.13	0.16	0.25	0.18	0.18	0.17	0.17	0.25
		<i>RIVAR</i>	0.15	0.10	0.06	0.08	0.15	0.24	0.15	0.11	0.14	0.27
0.6	0.6	<i>RIMSE</i>	0.35	0.26	0.13	0.25	0.40	0.38	0.31	0.15	0.25	0.40
		<i>RIBIAS</i> ²	0.34	0.26	0.12	0.24	0.39	0.36	0.29	0.13	0.23	0.38
		<i>RIVAR</i>	0.07	0.06	0.05	0.06	0.08	0.12	0.10	0.07	0.10	0.15
INFEASIBLE		<i>RIMSE</i>	0.13	0.11	0.09	0.10	0.11	0.13	0.11	0.09	0.10	0.12

Table 2. Design I, $n = 10000$, $T = 2$. Monte Carlo results of estimation of functions $\hat{h}(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. *RIMSE*, *RIBIAS*², and *RIVAR* stand for square roots of integrated Mean Squared Error, Squared Bias, and Variance, respectively.

DESIGN II, $n = 2500, T = 2$												
h_Y	h_w		$\varepsilon_{it} \sim Laplace$					$\varepsilon_{it} \sim N(0, 1)$				
			0.875	0.75	0.5	0.25	0.125	0.875	0.75	0.5	0.25	0.125
0.4	0.4	<i>RIMSE</i>	0.46	0.24	0.16	0.34	0.69	0.48	0.26	0.22	0.36	0.59
		<i>RIBIAS</i> ²	0.35	0.15	0.09	0.22	0.49	0.31	0.11	0.09	0.28	0.51
		<i>RIVAR</i>	0.30	0.18	0.14	0.26	0.48	0.37	0.24	0.20	0.22	0.30
0.4	0.6	<i>RIMSE</i>	0.28	0.27	0.13	0.31	0.56	0.33	0.32	0.16	0.41	0.68
		<i>RIBIAS</i> ²	0.19	0.23	0.07	0.28	0.53	0.20	0.25	0.07	0.37	0.65
		<i>RIVAR</i>	0.20	0.14	0.11	0.13	0.18	0.26	0.19	0.15	0.16	0.20
0.6	0.4	<i>RIMSE</i>	0.47	0.25	0.18	0.32	0.60	0.53	0.29	0.24	0.40	0.60
		<i>RIBIAS</i> ²	0.37	0.20	0.14	0.23	0.47	0.37	0.18	0.17	0.33	0.53
		<i>RIVAR</i>	0.29	0.16	0.12	0.21	0.38	0.38	0.22	0.17	0.22	0.28
0.6	0.6	<i>RIMSE</i>	0.31	0.27	0.14	0.33	0.56	0.34	0.29	0.17	0.42	0.67
		<i>RIBIAS</i> ²	0.26	0.24	0.10	0.31	0.54	0.25	0.25	0.11	0.40	0.65
		<i>RIVAR</i>	0.17	0.12	0.09	0.10	0.13	0.24	0.16	0.12	0.14	0.17
INFEASIBLE		<i>RIMSE</i>	0.24	0.17	0.14	0.17	0.22	0.24	0.17	0.14	0.16	0.22

Table 3. Design II, $n = 2500, T = 2$. Monte Carlo results of estimation of functions $\hat{h}(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. *RIMSE*, *RIBIAS*², and *RIVAR* stand for square roots of integrated Mean Squared Error, Squared Bias, and Variance, respectively.

DESIGN II, $n = 10000, T = 2$												
h_Y	h_w		$\varepsilon_{it} \sim Laplace$					$\varepsilon_{it} \sim N(0, 1)$				
			0.875	0.75	0.5	0.25	0.125	0.875	0.75	0.5	0.25	0.125
0.4	0.4	<i>RIMSE</i>	0.38	0.19	0.11	0.26	0.52	0.41	0.20	0.15	0.31	0.53
		<i>RIBIAS</i> ²	0.35	0.16	0.08	0.19	0.42	0.33	0.13	0.11	0.27	0.48
		<i>RIVAR</i>	0.16	0.10	0.07	0.17	0.31	0.23	0.15	0.11	0.15	0.22
0.4	0.6	<i>RIMSE</i>	0.24	0.24	0.09	0.29	0.52	0.26	0.26	0.10	0.36	0.62
		<i>RIBIAS</i> ²	0.21	0.23	0.07	0.28	0.51	0.21	0.23	0.06	0.35	0.60
		<i>RIVAR</i>	0.11	0.07	0.05	0.06	0.08	0.16	0.11	0.08	0.10	0.12
0.6	0.4	<i>RIMSE</i>	0.42	0.22	0.15	0.24	0.49	0.48	0.25	0.20	0.33	0.53
		<i>RIBIAS</i> ²	0.39	0.20	0.14	0.20	0.42	0.40	0.20	0.17	0.30	0.49
		<i>RIVAR</i>	0.15	0.08	0.06	0.12	0.24	0.25	0.14	0.10	0.14	0.21
0.6	0.6	<i>RIMSE</i>	0.28	0.25	0.11	0.31	0.53	0.29	0.25	0.13	0.38	0.62
		<i>RIBIAS</i> ²	0.26	0.24	0.11	0.31	0.53	0.25	0.23	0.11	0.37	0.61
		<i>RIVAR</i>	0.09	0.06	0.05	0.05	0.06	0.14	0.09	0.07	0.08	0.10
INFEASIBLE		<i>RIMSE</i>	0.15	0.10	0.09	0.10	0.13	0.14	0.10	0.09	0.10	0.13

Table 4. Design II, $n = 10000, T = 2$. Monte Carlo results of estimation of functions $\hat{h}(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. *RIMSE*, *RIBIAS*², and *RIVAR* stand for square roots of integrated Mean Squared Error, Squared Bias, and Variance, respectively.

DESIGN III, $n = 2500$, $T = 2$

h_Y	h_w		$\varepsilon_{it} \sim Laplace$					$\varepsilon_{it} \sim N(0, 1)$				
			0.875	0.75	0.5	0.25	0.125	0.875	0.75	0.5	0.25	0.125
0.4	0.4	<i>RIMSE</i>	0.29	0.23	0.19	0.23	0.44	0.37	0.29	0.23	0.25	0.35
		<i>RIBIAS</i> ²	0.16	0.14	0.10	0.07	0.24	0.20	0.17	0.10	0.08	0.18
		<i>RIVAR</i>	0.25	0.18	0.16	0.22	0.37	0.31	0.24	0.21	0.24	0.30
0.6	0.4	<i>RIMSE</i>	0.40	0.17	0.14	0.18	0.37	0.44	0.35	0.26	0.24	0.35
		<i>RIBIAS</i> ²	0.36	0.11	0.07	0.11	0.33	0.27	0.26	0.17	0.09	0.19
		<i>RIVAR</i>	0.17	0.14	0.12	0.14	0.16	0.35	0.23	0.19	0.23	0.29
0.4	0.6	<i>RIMSE</i>	0.36	0.26	0.21	0.20	0.39	0.44	0.21	0.17	0.22	0.43
		<i>RIBIAS</i> ²	0.24	0.21	0.15	0.07	0.23	0.37	0.11	0.08	0.12	0.37
		<i>RIVAR</i>	0.27	0.16	0.14	0.18	0.32	0.25	0.19	0.16	0.18	0.23
0.6	0.6	<i>RIMSE</i>	0.41	0.20	0.16	0.18	0.37	0.42	0.23	0.20	0.22	0.42
		<i>RIBIAS</i> ²	0.39	0.16	0.13	0.14	0.34	0.35	0.16	0.13	0.14	0.37
		<i>RIVAR</i>	0.14	0.12	0.1	0.11	0.13	0.23	0.17	0.14	0.16	0.19
INFEASIBLE		<i>RIMSE</i>	0.22	0.19	0.20	0.18	0.24	0.21	0.20	0.20	0.18	0.24

Table 5. Design III, $n = 2500$, $T = 2$. Monte Carlo results of estimation of functions $\hat{h}(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. *RIMSE*, *RIBIAS*², and *RIVAR* stand for square roots of integrated Mean Squared Error, Squared Bias, and Variance, respectively.

DESIGN III, $n = 10000$, $T = 2$

h_Y	h_w		$\varepsilon_{it} \sim Laplace$					$\varepsilon_{it} \sim N(0, 1)$				
			0.875	0.75	0.5	0.25	0.125	0.875	0.75	0.5	0.25	0.125
0.4	0.4	<i>RIMSE</i>	0.18	0.17	0.13	0.14	0.33	0.25	0.23	0.16	0.18	0.26
		<i>RIBIAS</i> ²	0.12	0.14	0.10	0.06	0.23	0.13	0.18	0.10	0.09	0.14
		<i>RIVAR</i>	0.13	0.09	0.09	0.13	0.24	0.21	0.15	0.12	0.16	0.21
0.4	0.6	<i>RIMSE</i>	0.37	0.13	0.09	0.13	0.32	0.37	0.15	0.11	0.16	0.37
		<i>RIBIAS</i> ²	0.36	0.11	0.07	0.11	0.31	0.35	0.10	0.07	0.11	0.34
		<i>RIVAR</i>	0.08	0.07	0.06	0.07	0.08	0.14	0.11	0.09	0.11	0.14
0.6	0.4	<i>RIMSE</i>	0.28	0.23	0.18	0.13	0.29	0.34	0.30	0.21	0.17	0.25
		<i>RIBIAS</i> ²	0.24	0.21	0.16	0.08	0.23	0.24	0.27	0.18	0.09	0.17
		<i>RIVAR</i>	0.15	0.09	0.08	0.10	0.18	0.25	0.13	0.11	0.14	0.18
0.6	0.6	<i>RIMSE</i>	0.40	0.17	0.13	0.14	0.34	0.39	0.18	0.16	0.17	0.37
		<i>RIBIAS</i> ²	0.39	0.16	0.12	0.13	0.33	0.36	0.16	0.14	0.14	0.36
		<i>RIVAR</i>	0.08	0.06	0.05	0.05	0.06	0.14	0.09	0.08	0.09	0.11
INFEASIBLE		<i>RIMSE</i>	0.13	0.12	0.11	0.10	0.14	0.13	0.12	0.11	0.10	0.14

Table 6. Design III, $n = 10000$, $T = 2$. Monte Carlo results of estimation of functions $h(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. *RIMSE*, *RIBIAS*², and *RIVAR* stand for square roots of integrated Mean Squared Error, Squared Bias, and Variance, respectively.

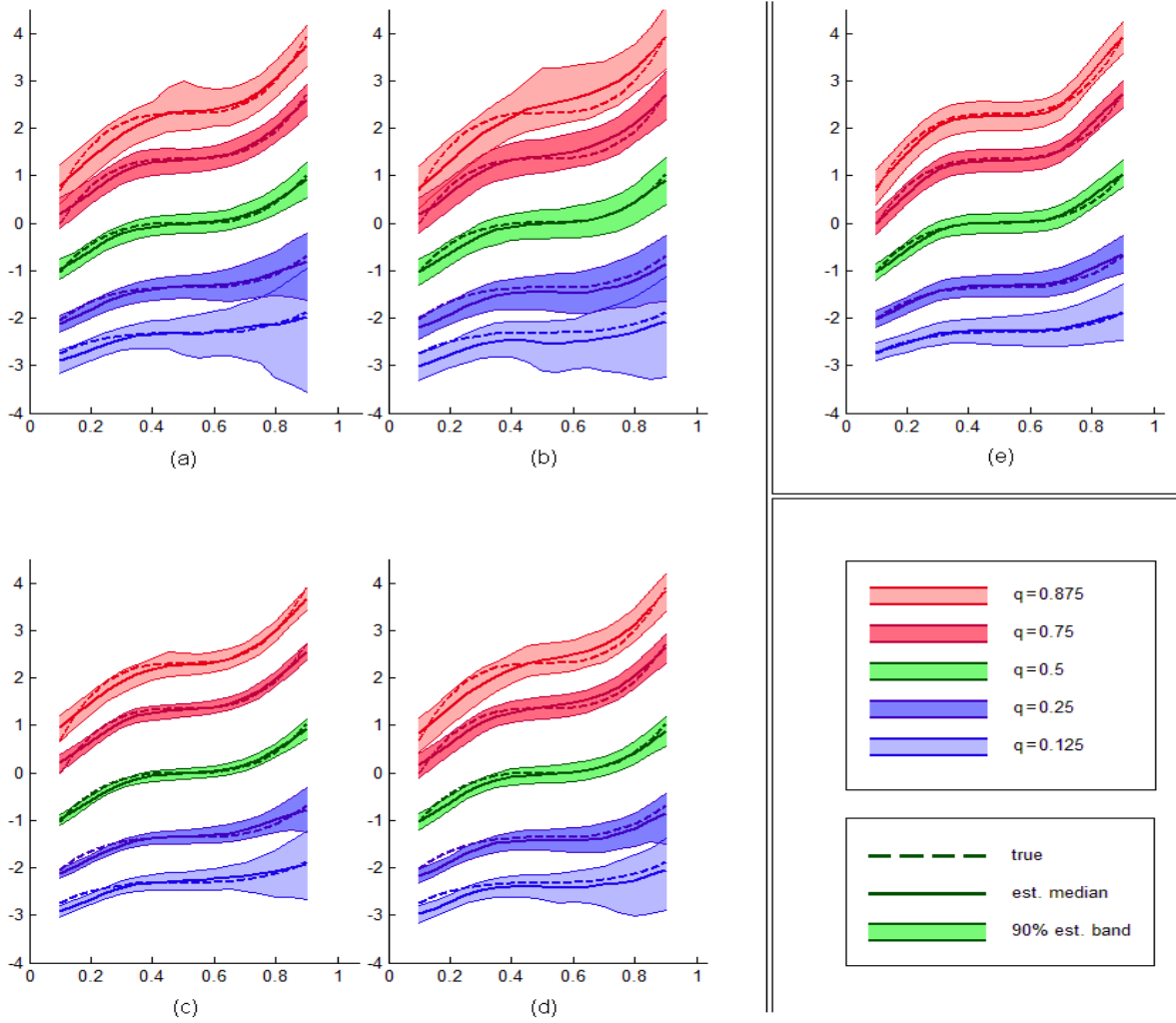


Figure 1. Design I. Monte Carlo results of estimation of functions $h(\cdot, q)$ for quantiles $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$. Panels (a) and (b) correspond to $n = 2500$, while panels (c) and (d) correspond to $n = 10000$. Panels (a) and (c) correspond to $\varepsilon_{it} \sim i.i.d. Laplace$, and panels (b) and (d) correspond to $\varepsilon_{it} \sim i.i.d. N(0, 1)$. The bandwidths are $h_Y = h_w = 0.4$ for panels (a)-(c). Panel (e) presents the infeasible estimator (see the main text for description) with $n = 2500$ and $\varepsilon_{it} \sim i.i.d. N(0, 1)$. For each $x \in [0.1, 0.9]$ and $q \in \{0.875, 0.75, 0.5, 0.25, 0.125\}$ the true line corresponds to the true value of $h(x, q)$, estimated median line is the median of $\hat{h}(x, q)$ over the simulation runs, and estimation band represents the area between 0.05-th and 0.95-th quantiles of $\hat{h}(x, q)$ over the simulation runs.

References

- ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings,” *Econometrica*, 70(1), 91–117.
- ALTONJI, J. G., AND R. L. MATZKIN (2005): “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” *Econometrica*, 73(4), 1053–1102.
- ARELLANO, M., AND S. BONHOMME (2008): “Identifying Distributional Characteristics in Random Coefficients Panel Data Models,” Working Paper, CEMFI.
- ARELLANO, M., AND J. HAHN (2006): “Understanding Bias in Nonlinear Panel Models: Some Recent Developments,” in *Advances in Economics and Econometrics, Ninth World Congress*, ed. by R. Blundell, W. K. Newey, and T. Persson. Cambridge University Press.
- ATHEY, S., AND G. W. IMBENS (2006): “Identification and Inference in Nonlinear Difference-in-Differences Models,” *Econometrica*, 74(2), 431–497.
- BESTER, A., AND C. HANSEN (2007): “Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model,” Manuscript, University of Chicago.
- BHATTACHARYA, P. K., AND A. K. GANGOPADHYAY (1990): “Kernel and Nearest-Neighbor Estimation of a Conditional Quantile,” *The Annals of Statistics*, 18(3), 1400–1415.
- BILLINGSLEY, P. (1986): *Probability and Measure*. John Wiley and Sons, 2nd edn.
- BITLER, M. P., J. B. GELBACH, AND H. W. HOYNES (2006): “What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments,” *American Economic Review*, 96(4), 988–1012.
- BONHOMME, S., AND J.-M. ROBIN (2008): “Generalized nonparametric deconvolution with an application to earnings dynamics,” CeMMAP working papers CWP03/08, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- CARD, D. (1996): “The Effect of Unions on the Structure of Wages: A Longitudinal Analysis,” *Econometrica*, 64(4), 957–79.
- CARD, D., T. LEMIEUX, AND W. C. RIDDELL (2004): “Unions and Wage Inequality,” *Journal of Labor Research*, 25(4), 519–562.
- CHAMBERLAIN, G. (1992): “Efficiency Bounds for Semiparametric Regression,” *Econometrica*, 60(3), 567–596.
- CHEN, X., AND M. REISS (2007): “On Rate Optimality for Ill-posed Inverse Problems in Econometrics,” Cowles Foundation Discussion Papers 1626, Cowles Foundation, Yale University.
- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, AND A. GALICHON (2007): “Quantile and probability curves without crossing,” CeMMAP working papers CWP10/07, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, J. HAHN, AND W. NEWEY (2008): “Identification and estimation of marginal effects in nonlinear panel models,” CeMMAP working papers CWP25/08, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, AND W. K. NEWEY (2009): “Quantile and Average Effects in Nonseparable Panel Models,” Working Paper, MIT.
- CHERNOZHUKOV, V., AND C. HANSEN (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73(1), 245–261.
- CHESHER, A. (2003): “Identification in Nonseparable Models,” *Econometrica*, 71(5), 1405–1441.
- DELAIGLE, A., AND P. HALL (2006): “On optimal kernel choice for deconvolution,” *Statistics & Probability Letters*, 76(15), 1594–1602.
- DELAIGLE, A., P. HALL, AND A. MEISTER (2008): “On deconvolution with repeated measurements,” *Annals of Statistics*, 36(02), 665–685.
- DJEBBARI, H., AND J. SMITH (2008): “Heterogeneous impacts in PROGRESA,” *Journal of Econometrics*, 145(1-2), 64–80.
- EVDOKIMOV, K. (2009): “Further Results on Identification of Panel Data Models with Unobserved Heterogeneity,” Working Paper, Yale University.
- (2010): “Essays on Nonparametric and Semiparametric Econometric Models,” Ph.D. thesis, Yale University.
- FAN, J. (1991): “On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems,” *Annals of Statistics*, 19(03), 1257–1272.
- GIL-PELAEZ, J. (1951): “Note on the Inversion Theorem,” *Biometrika*, 38, 481–482.
- GRAHAM, B., J. HAHN, AND J. POWELL (2009): “A Quantile Panel Data Regression Model,” Working Paper, UC Berkeley.
- GRAHAM, B., AND J. POWELL (2008): “Identification and Estimation of Irregular Correlated Random Coefficient Models,” Working Paper 14469, National Bureau of Economic Research.
- HAHN, J., AND W. NEWEY (2004): “Jackknife and Analytical Bias Reduction for Nonlinear Panel Models,” *Econometrica*, 72(4), 1295–1319.
- HALL, P., AND S. LAHIRI (2008): “Estimation of distributions, moments and quantiles in deconvolution problems,” *Annals of Statistics*, 36(5), 2110–2134.
- HAYASHI, F., J. ALTONJI, AND L. KOTLIKOFF (1996): “Risk-Sharing between and within Families,” *Econometrica*, 64(2), 261–94.
- HECKMAN, J., AND E. VYTLACIL (1998): “Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling,” *The Journal of Human Resources*, 33(4), 974–987.

- HECKMAN, J. J., AND T. E. MACURDY (1980): "A Life Cycle Model of Female Labour Supply," *Review of Economic Studies*, 47(1), 47–74.
- HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64(4), 487–535.
- HENDERSON, D. J., R. J. CARROLL, AND Q. LI (2008): "Nonparametric estimation and testing of fixed effects panel data models," *Journal of Econometrics*, 144(1), 257–275.
- HODERLEIN, S., AND H. WHITE (2009): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," Working Paper, Brown University.
- HONORÉ, B. E., AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74(3), 611–629.
- HOROWITZ, J. L., AND M. MARKATOU (1996): "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economic Studies*, 63(1), 145–168.
- HSIAO, C. (2003): *Analysis of Panel Data*. Cambridge University Press.
- HSIAO, C., AND M. H. PESARAN (2004): "Random Coefficient Panel Data Models," IZA Discussion Papers 1236, Institute for the Study of Labor (IZA).
- HU, Y., AND G. RIDDER (2005): "Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information," IEPR Working Papers 05.39, Institute of Economic Policy Research (IEPR).
- IMBENS, G., AND W. NEWEY (In press): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica*.
- KITAMURA, Y. (2004): "Nonparametric Identifiability Of Finite Mixtures," Manuscript, Department of Economics, Yale University.
- KOTLARSKI, I. (1967): "On Characterizing The Gamma And The Normal Distribution," *Pacific Journal Of Mathematics*, 20(1), 69–76.
- LEMIEUX, T. (1998): "Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection," *Journal of Labor Economics*, 16(2), 261–291.
- LI, T., AND Q. VUONG (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.
- MACURDY, T. E. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting," *The Journal of Political Economy*, 89(6), 1059–1085.
- MATZKIN, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.
- MÜLLER, H.-G. (1991): "Smooth optimum kernel estimators near endpoints," *Biometrika*, 78(3), 521–530.

- NEWKEY, W. K. (1994): “Kernel Estimation of Partial Means and a General Variance Estimator,” *Econometric Theory*, 10, 233–253.
- OULD-SAÏD, E., D. YAHIA, AND A. NECIR (2009): “A strong uniform convergence rate of a kernel conditional quantile estimator under random left-truncation and dependent data,” *Electronic Journal of Statistics*, 3, 426–445.
- PORTER, J. (1996): “Essays in Econometrics,” Ph.D. thesis, MIT.
- SCHENNACH, S. M. (2004a): “Estimation of Nonlinear Models with Measurement Error,” *Econometrica*, 72(1), 33–75.
- (2004b): “Nonparametric Estimation in the Presence of Measurement Error,” *Econometric Theory*, 20, 1046–1093.
- STEFANSKI, L., AND R. CARROLL (1990): “Deconvoluting kernel density estimators,” *Statistics*, 21, 169–184.
- STONE, C. J. (1982): “Optimal Global Rates of Convergence for Nonparametric Regression,” *The Annals of Statistics*, 10(4), 1040–1053.
- WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.