Econometric Theory, **0**, 2026, 1–30.
doi:10.1017/S0266466626100589

IDENTIFICATION AND ESTIMATION OF A NONPARAMETRIC TIME-VARYING PANEL DATA MODEL WITH COMPLETELY MISSING REGRESSORS IN SOME PERIODS

JIANGANG ZENG
Fordham University

This article considers a nonparametric time-varying panel data model in which the primary regressor of interest is completely missing in some periods. Under standard assumptions, identification results are established for the unknown functions and the distributions of the missing regressor. Specifically, one period of completely observed data suffices to identify the random effects model, whereas two periods are required for the fixed effects model. Following identification, conditional deconvolution kernel estimators are developed, and their convergence rates are derived. Simulation results show that these estimators perform well in small samples, and an empirical exercise demonstrates the relevance and practical potential of the proposed method.

1. INTRODUCTION

The panel data model has received considerable attention for its ability to account for unobserved persistent heterogeneity and mitigate omitted variable bias (Wooldridge, 2010). Concurrently, nonparametric regression has gained popularity due to its flexibility in modeling complex relationships (Li and Racine, 2007). Recent research integrates these strengths into the following nonparametric time-varying panel data model (Porter, 1996; Wang, 2003; Henderson, Carroll, and Li, 2008):

$$Y_{it} = f_t(X_{it}, \mathbf{Z}_{it}) + \alpha_i + \eta_{it}, \qquad i = 1, \ldots, n, \quad t = 1, \ldots, T. \tag{1}$$

In this specification, Y_{it} is a scalar outcome, X_{it} is the primary regressor of interest, and $\mathbf{Z}_{it}$ is a vector of controls. The term α_i denotes unobserved persistent heterogeneity, potentially correlated with the regressors, and η_{it} is an idiosyncratic

I thank Jason Abrevaya, Brendan Kline, Le Wang, and the seminar participants at the 2022 Midwest Econometrics Group for helpful discussions. I also thank the Editors, Peter C. B. Phillips and Liangjun Su, the Co-Editor, Iván Fernández-Val, and two anonymous referees for constructive comments and suggestions that greatly improved the article. All errors are my own. Address correspondence to Jiangang Zeng, Fordham University, USA, e-mail: jzeng59@fordham.edu.

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

error term. The unknown time-varying function $f_t(\cdot)$ captures the relationship between the regressors and the outcome.

Various methods have been developed to identify and estimate the unknown function $f_t(x_t, \mathbf{z}_t)$ in Equation (1). However, in empirical applications, the primary regressor X_{it} may be entirely unavailable in some periods. Herein, we define X_{it} as "completely missing" if it is unobserved for all i in a given period t, and "completely observed" if it is observed for all i. To align with the literature, we use "missing" to denote cases, where X_{it} is unobserved for only a subset of units in period t. These definitions apply analogously to all other variables and vectors in the model.

Several factors can cause X_{it} to be completely missing in some periods while Y_{it} and $\mathbf{Z}_{it}$ remain completely observed. First, budgetary, legal, or temporal constraints may impede the collection of X_{it}; similarly, limited storage capacity or varying data retention policies may result in its loss over time. In such cases, the outcome variable and other regressors remain available because they are unaffected by these constraints or can be obtained from alternative sources. Second, when merging datasets, the temporal coverage of X_{it} may span only a subset of the periods available for Y_{it} and $\mathbf{Z}_{it}$. We illustrate this scenario by examining the effect of mathematics vocabulary (X_{it}) on computation (Y_{it}) among fourth- and sixth-grade Chinese elementary students. Due to funding and time constraints, mathematics vocabulary (MV) data are collected only for the fourth grade. However, computation data and student characteristics ($\mathbf{Z}_{it}$), such as socioeconomic status (SES) and age, are available for both grades through separate testing and administrative records.

Missing data present substantial identification and estimation challenges in econometrics (Chaudhuri and Guilkey, 2016; Abrevaya and Donald, 2017). These challenges are exacerbated when X_{it} is completely absent in some periods—particularly when $f_t(\cdot)$ is time-varying—often leading to inaccurate or incomplete analysis.

In this article, we propose a novel approach to identify and estimate $f_t(x_t, \mathbf{z}_t)$ for all $(x_t, \mathbf{z}_t)$ on the support, across all time periods, allowing for X_{it} to be completely missing in some periods. Moreover, the distributions of the missing X_{it} can also be identified and estimated. Our analysis covers the random (i.e., α_i is independent of the regressors) and the fixed (i.e., α_i is correlated with the regressors) effects models.

For the random effects model, we first identify $f_t(x_t, \mathbf{z}_t)$ and the distributions of $\alpha_i + \eta_{it}$ in periods where X_{it} is completely observed (Wang, 2003). Using the time-homogeneity assumption that the distribution of η_{it} is invariant over time, we then recover the distributions of $\alpha_i + \eta_{it}$ in periods with completely missing X_{it}. This assumption, widely adopted in the panel data literature, can be interpreted as "time is randomly assigned" or "time is an instrument" (Abrevaya, 1999; Chernozhukov et al., 2013). Next, under appropriate independence conditions, we deconvolve the distribution of $\alpha_i + \eta_{it}$ from the outcome variable to recover the conditional distributions of the $f_t(X_{it}, \mathbf{Z}_{it})$ given $\mathbf{Z}_{it}$. These conditional distributions, combined

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

with nonseparable model identification strategies (Matzkin, 2003; Altonji and Matzkin, 2005), enable identification of $f_t(x_t, \mathbf{z}_t)$ in periods with missing data and the distributions of the missing X_{it}. Because α_i and $(X_{it}, \mathbf{Z}_{it})$ are independent, one period of completely observed data suffices for identification.

Identification of $f_t(x_t, \mathbf{z}_t)$ is straightforward when X_{it} is fully observed. When X_{it} is missing, identification relies on using fully observed periods to "impute" the conditional distributions of $f_t(X_{it}, \mathbf{Z}_{it})$. This imputation is grounded in the time-homogeneity assumption and implemented via conditional deconvolution.

We develop estimators for $f_t(x_t, \mathbf{z}_t)$ that follow the identification strategy. When X_{it} is fully observed, $f_t(x_t, \mathbf{z}_t)$ is a conditional expectation estimated using the standard Nadaraya–Watson (NW) kernel estimator (Li and Racine, 2007, Chap. 2). When X_{it} is completely missing, we express $f_t(x_t, \mathbf{z}_t)$ as a combination of a conditional cumulative distribution function (CDF) and a conditional quantile. Specifically, we estimate the conditional CDF via conditional deconvolution using NW kernel estimators for the required conditional characteristic functions and obtain the conditional quantile by inverting the estimated CDF. Under smoothness and regularity conditions on the random variables and functions, along with standard requirements on kernels and bandwidths, we establish uniform convergence rates for these estimators (Fan, 1991b; Evdokimov, 2010). Simulation results show strong performance in small samples.

The key distinction between fixed and random effects models is that the former allows for correlation between the unobserved persistent heterogeneity α_i and the regressors $(X_{it}, \mathbf{Z}_{it})$. We address this endogeneity by first-differencing to eliminate α_i, after which identification proceeds similarly to the random effects case. When X_{it} is fully observed, we apply standard nonparametric identification; when it is completely missing, we deconvolve the error terms—recovered under time-homogeneity—from the outcome variables. Since first-differencing is used, identification requires two periods of completely observed data, consistent with the fixed effects literature (Wooldridge, 2010, Chap. 10).

We construct estimators for $f_t(x_t, \mathbf{z}_t)$ in the fixed effects framework that closely mirror those for the random effects model but include additional conditioning variables due to first-differencing. Under appropriate assumptions, we establish uniform convergence rates for these estimators. Simulation results again indicate strong small-sample performance.

To illustrate the empirical relevance of the proposed method, we examine the effect of MV on computation among fourth- and sixth-grade students in China. As noted, MV data for sixth-grade students are completely missing, while all other variables remain fully observed. The results indicate that MV significantly affects computation in both grades.

This article contributes to the missing data literature in three ways. First, while economists typically study models in which regressors are missing, this article focuses on settings where a regressor is entirely absent in some periods—a common challenge in empirical research. Because the variable is completely missing, the proposed approach eliminates the need for standard assumptions

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

about missingness mechanisms, such as missing completely at random, missing at random, or missing not at random. Second, this article examines missing data within a panel data framework, a topic that has received limited attention. Existing studies have primarily considered parametric and semiparametric panel data models (Moffit, Fitzgerald, and Gottschalk, 1999; Chen, Yi, and Cook, 2010; Abrevaya, 2019; Coe, 2019), whereas this study develops a nonparametric time-varying panel data model designed to capture nonlinear relationships. The framework also incorporates unobserved persistent heterogeneity (α_i), a critical feature often overlooked in prior work. Third, estimation relies on conditional deconvolution and kernel methods. To our knowledge, this is the first application of these techniques to models with missing data, in contrast to standard generalized method of moments, imputation, or dummy-variable approaches (Chaudhuri and Guilkey, 2016; Abrevaya and Donald, 2017; Muris, 2020).

This article also relates to the measurement error literature, in which the variable of interest is mismeasured and therefore unobserved—a setting conceptually similar to complete missingness. A closely related study, Wilhelm (2015), examines the model $Y_{it} = f(X_{it}^{\star}) + \alpha_i + \eta_{it}$, where the true regressor $X_{it}^{\star}$ is unobserved but a mismeasured proxy $X_{it} = X_{it}^{\star} + U_{it}$ is fully observed. Our setup and identification strategy differ in several respects. First, while Wilhelm (2015) requires a mismeasured but fully observed variable X_{it}, our approach does not. Second, rather than relying on instrumental variables (e.g., lagged regressors or outcomes) and their associated independence assumptions, we exploit the time-homogeneity of the error terms. Third, their unknown function is time-invariant, whereas ours is time-varying. Finally, Wilhelm (2015) allows $X_{it}^{\star}$ to be unobserved in all periods, whereas our framework requires at least some periods with complete observations. Other significant identification results in this literature, such as Hu and Schennach (2008), Schennach (2013), and Hu and Sasaki (2014), depend on instrumental variables, repeated measurements, or auxiliary data, none of which are required in our setting.

The remainder of this article is organized as follows. Section 2 establishes identification for the random and fixed effects models. Section 3 develops the corresponding estimators and derives their convergence rates. Section 4 evaluates these estimators through simulation, Section 5 presents an empirical exercise, and Section 6 concludes.

2. IDENTIFICATION

This section presents the identification results. We first show that, for the random effects model, one period of completely observed data suffices to identify $f_t(x_t, \mathbf{z}_t)$ for all $(x_t, \mathbf{z}_t)$ and t. We then establish that the fixed effects model requires two such periods. We also show that the distributions of the completely missing X_{it} are identified in both frameworks.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

2.1. Random Effects Model

We consider the identification of the following nonparametric random effects model for $T = 2$ ($t = 1, 2$) periods, where X_{it} is completely missing in one period:

$$Y_{it} = f_t(X_{it}, \mathbf{Z}_{it}) + \alpha_i + \eta_{it} = f_t(X_{it}, \mathbf{Z}_{it}) + \epsilon_{it}.$$

In this model, Y_{it}, X_{it}, α_i, and η_{it} are scalars, and $\mathbf{Z}_{it}$ is a p-dimensional vector. Given the model's static structure, we assume without loss of generality that X_{i1} is entirely missing, while $(X_{i2}, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, Y_{i1}, Y_{i2})$ are fully observed. The identification argument remains unchanged if X_{i2} is the missing variable. Extending the framework to $T > 2$ periods with τ periods of missing X_{it}, where $\tau \in \{1, \ldots, T-1\}$, is straightforward. Identification relies on the following assumptions.

Assumption RE1 (Independence). (i) α_i and (η_{i1}, η_{i2}) are independent. (ii) ϵ_{it} and $(X_{it}, \mathbf{Z}_{it})$ are independent for $t \in \{1, 2\}$. (iii) $\mathbf{E}[\epsilon_{i2}] = 0$. (iv) X_{i1} and $\mathbf{Z}_{i1}$ are independent.

Assumption RE1 imposes several independence conditions. Assumption RE1(i) is standard for random effects panel data models. Assumption RE1(ii) imposes contemporaneous independence, which is weaker than strict exogeneity because it does not require ϵ_{i1} or ϵ_{i2} to be independent of $(X_{i1}, \mathbf{Z}_{i1}, X_{i2}, \mathbf{Z}_{i2})$ (Porter, 1996; Wooldridge, 2010, Chap. 10). Assumption RE1(iii) serves as a normalization. Assumption RE1(iv) rules out dependence between X_{i1} and $\mathbf{Z}_{i1}$, which may be restrictive; we discuss potential remedies following Theorem 1. Notably, Assumption RE1 implies that X_{i1} and ϵ_{i1} are independent conditional on $\mathbf{Z}_{i1}$, a property essential for applying conditional deconvolution.

Assumption RE2 (Time-Homogeneity). The marginal distributions of η_{i1} and η_{i2} are the same.

The time-homogeneity assumption is widely used in the panel data literature (Abrevaya, 1999; Chernozhukov et al., 2013). It requires that the marginal distribution of η_{it} does not vary with t and can be interpreted as "time is randomly assigned" or "time is an instrument" (Chernozhukov et al., 2013). Assumptions RE1(i) and RE2 together imply that the marginal distributions of ϵ_{i1} and ϵ_{i2} are identical. This equivalence allows the recovery of the distribution of $f_1(X_{i1}, \mathbf{Z}_{i1})$ (the period with completely missing data) using information from the fully observed second period. Additional details are provided in the proof of Theorem 1.

Assumption RE3 (Properties of $f_1(X_{i1}, \mathbf{Z}_{i1})$). (i) For each $\mathbf{z}_1$ on the support, $x_1 \to f_1(x_1, \mathbf{z}_1)$ is strictly monotone on the support of X_{i1}. (ii) For some given $\bar{\mathbf{z}}_1$, $f_1(x_1, \bar{\mathbf{z}}_1) = x_1$ for all x_1 on the support of X_{i1}.

Assumption RE3(i) requires $f_1(X_{i1}, \mathbf{z}_1)$ to be strictly monotonic in X_{i1} for all $\mathbf{z}_1$ on the support. This condition is standard in the nonseparable model literature (Chesher, 2003; Matzkin, 2003) and ensures invertibility of $f_1(X_{i1}, \mathbf{Z}_{i1})$ with respect to X_{i1}, thereby enabling inference from observed variation. In this article,

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

we further assume that $f_1(X_{i1}, \mathbf{z}_1)$ is strictly increasing, a property satisfied in many economic applications. For example, in the wage-education framework, higher education is associated with higher wages. Similarly, in our empirical analysis, greater MV is expected to improve computation accuracy. Many common functional forms, including linear and multiplicative specifications, satisfy this condition by construction.

Assumption RE3(ii) is a standard normalization (Matzkin, 2003; Altonji and Matzkin, 2005). It fixes the value of $f_1(\cdot)$ at some specific $\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1$. Alternative normalization strategies include specifying the distribution of X_{i1} or assuming that $f_1(\cdot)$ is homogeneous (Matzkin, 2003; Evdokimov, 2010).

Assumption RE4. (i) The characteristic function of ϵ_{it}, denoted by $\phi_{\epsilon_{it}}(s)$, does not vanish for any $s \in \mathbb{R}$ and $t \in \{1, 2\}$. (ii) $\mathbf{E}[|f_1(X_{i1}, \mathbf{Z}_{i1})||\mathbf{Z}_{i1}]$ is uniformly bounded almost surely.

Assumption RE4 is a standard, mild condition ensuring the existence of the conditional characteristic functions of ϵ_{it} and $f_1(X_{i1}, \mathbf{Z}_{i1})$. These functions are used to derive the conditional CDF of $f_1(X_{i1}, \mathbf{Z}_{i1})$, which enables the identification of $f_1(x_1, \mathbf{z}_1)$ and the distribution of X_{i1}.

THEOREM 1. *If Assumptions RE1–RE4 are satisfied, then $f_t(x_t, \mathbf{z}_t)$ is identified for all $(x_t, \mathbf{z}_t)$ on the support of $(X_{it}, \mathbf{Z}_{it})$, for $t \in \{1, 2\}$. Moreover, the distribution of X_{i1} is identified.*

Identifying $f_2(x_2, \mathbf{z}_2)$ is straightforward and standard (Wang, 2003), whereas identifying $f_1(x_1, \mathbf{z}_1)$ is more challenging because X_{i1} is completely missing. To address this, we use second-period data to "impute" first-period variation. Specifically, the fully observed second-period data identify the distribution of ϵ_{i2}, which, under time-homogeneity, is identical to that of ϵ_{i1}. Deconvolving ϵ_{i1} from Y_{i1} conditional on $\mathbf{Z}_{i1}$ yields the conditional distribution of $f_1(X_{i1}, \mathbf{Z}_{i1})$. This enables the identification of $f_1(x_1, \mathbf{z}_1)$ and the distribution of X_{i1} using standard nonseparable model techniques (Matzkin, 2003; Altonji and Matzkin, 2005).[1] The formal proof is provided in Appendix B.

As noted in the discussion of Assumption RE1, the independence between X_{i1} and $\mathbf{Z}_{i1}$ may be restrictive. One approach is to introduce control variables $\mathbf{W}_{i1}$ such that X_{i1} and $\mathbf{Z}_{i1}$ are independent conditional on $\mathbf{W}_{i1}$. Under additional technical conditions, $f_1(x_1, \mathbf{z}_1)$ can be identified as follows:

$$f_1(x_1, \mathbf{z}_1) = F^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}\Big(\big(F_{f_1(X_{i1}, \mathbf{Z}_{i1})}(x_1 | \mathbf{Z}_{i1} = \bar{\mathbf{z}}_1, \mathbf{W}_{i1} = \mathbf{w}_1)\big) \big| \mathbf{Z}_{i1} = \mathbf{z}_1, \mathbf{W}_{i1} = \mathbf{w}_1\Big).$$

[1] The identification strategy employed herein, as well as in prior work, such as Matzkin (2003) and Altonji and Matzkin (2005), bears some resemblance to Lewbel's special regressor approach (Lewbel, 2014). In both cases, identification relies on the distribution of the dependent variable conditional on selected regressors. However, because the dependent variable in our setting is continuous rather than limited, we do not require the additivity or large-support assumptions typical of the special regressor literature.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

An alternative approach uses time-homogeneity (Abrevaya, 1999; Chernozhukov et al., 2013). Because X_{i1} and X_{i2} represent the same variable across periods, it is reasonable to assume they are identically distributed (e.g., a stable wage distribution over time). Under this assumption, the fully observed X_{i2} can proxy for the missing X_{i1}, addressing dependence between X_{i1} and $\mathbf{Z}_{i1}$. This approach relies on the following assumptions.

Assumption RE1′ (Independence). (i) α_i and (η_{i1}, η_{i2}) are independent. (ii) ϵ_{it} and $(X_{it}, \mathbf{Z}_{i1}, \mathbf{Z}_{i2})$ are independent for $t \in \{1,2\}$. (iii) $\mathbf{E}[\epsilon_{i2}] = 0$.

Assumption RE2′ (Time-Homogeneity). (i) The marginal distributions of η_{i1} and η_{i2} are the same. (ii) The distributions of X_{i1} and X_{i2} conditional on $(\mathbf{Z}_{i1}, \mathbf{Z}_{i2})$ are the same.

Assumption RE4′. (i) The characteristic function of ϵ_{it}, denoted by $\phi_{\epsilon_{it}}$, does not vanish for any $s \in \mathbb{R}$ and $t \in \{1,2\}$. (ii) $\mathbf{E}[|f_1(X_{i1}, \mathbf{Z}_{i1})||\mathbf{Z}_{i1}, \mathbf{Z}_{i2}]$ is uniformly bounded almost surely.

Assumption RE1′(ii) extends Assumption RE1(ii), enabling the deconvolution of ϵ_{i1} from Y_{i1} conditional on $(\mathbf{Z}_{i1}, \mathbf{Z}_{i2})$. This condition is stronger than contemporaneous independence but weaker than strict exogeneity. Assumption RE2′(ii) posits that time is conditionally randomly assigned with respect to X_{it}, which justifies the use of the fully observed X_{i2} as a proxy for the missing X_{i1}. Assumption RE4′(ii) serves as a mild technical condition.

PROPOSITION 1. *If Assumptions RE1′, RE2′, RE3(i), and RE4′ are satisfied, then $f_t(x_t, \mathbf{z}_t)$ is identified for all $(x_t, \mathbf{z}_t)$ on the support of $(X_{it}, \mathbf{Z}_{it})$, for $t \in \{1,2\}$. Moreover, the distribution of X_{i1} is identified.*

The proof parallels that of Theorem 1 and is therefore omitted. In particular, $f_1(x_1, \mathbf{z}_1)$ is identified as

$$f_1(x_1, \mathbf{z}_1) = F^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}\Big(\big(F_{X_{i2}}(x_1|\mathbf{Z}_{i1} = \mathbf{z}_1, \mathbf{Z}_{i2} = \mathbf{z}_2)\big)\big|\mathbf{Z}_{i1} = \mathbf{z}_1, \mathbf{Z}_{i2} = \mathbf{z}_2\Big).$$

Importantly, the normalization of $f_1(x_1, \mathbf{z}_1)$ (i.e., Assumption RE3(ii)) is not required under time-homogeneity of X_{it}. This allows a more direct comparison between $f_1(\cdot)$ and $f_2(\cdot)$, as well as their associated functionals, such as the average marginal effect.

2.2. Fixed Effects Model

We consider the identification of the following nonparametric fixed effects model for $T = 3$ $(t = 1, 2, 3)$ periods, where X_{it} is completely missing in one period:

$$Y_{it} = f_t(X_{it}, \mathbf{Z}_{it}) + \alpha_i + \eta_{it}.$$

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

In this model, Y_{it}, X_{it}, α_i, and η_{it} are scalars, and $\mathbf{Z}_{it}$ is a p-dimensional vector. Given the model's static structure, we assume without loss of generality that X_{i1} is entirely missing, while $(X_{i2}, X_{i3}, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \mathbf{Z}_{i3}, Y_{i1}, Y_{i2}, Y_{i3})$ are fully observed. The identification argument remains unchanged if either X_{i2} or X_{i3} is the missing variable. Extending the framework to $T > 3$ periods with τ periods of missing X_{it}, where $\tau \in \{1, \ldots, T-2\}$, is straightforward. Identification relies on the following assumptions, which are slight modifications of those for the random effects model.

Assumption FE1 (Independence). (i) $(\{\eta_{i\kappa}\}_{\kappa=1}^{3})$ are mutually independent and independent of $(\{X_{i\kappa}\}_{\kappa=1}^{3}, \{\mathbf{Z}_{i\kappa}\}_{\kappa=1}^{3})$. (ii) $\mathbf{E}[\eta_{i2}] = 0$. (iii) X_{i1} and $\mathbf{Z}_{i1}$ are independent.

Assumption FE2 (Time-Homogeneity). The marginal distributions of η_{i1}, η_{i2}, and η_{i3} are the same.

Assumption FE3 (Properties of $f_t(X_{it}, \mathbf{Z}_{it})$). (i) For each $\mathbf{z}_1$ on the support, $x_1 \to f_1(x_1, \mathbf{z}_1)$ is strictly monotone on the support of X_{i1}. (ii) For some given $\bar{\mathbf{z}}_1$, $f_1(x_1, \bar{\mathbf{z}}_1) = x_1$ for all x_1 on the support of X_{i1}. (iii) For some given $(\bar{x}_3, \bar{\mathbf{z}}_3)$, $f_3(\bar{x}_3, \bar{\mathbf{z}}_3) = 0$.

Assumption FE4. (i) The characteristic functions of $(\eta_{i1} - \eta_{i2})$ and $(\eta_{i3} - \eta_{i2})$, denoted by $\phi_{(\eta_{i1}-\eta_{i2})}(s)$ and $\phi_{(\eta_{i3}-\eta_{i2})}(s)$, do not vanish for any s $\in \mathbb{R}$. (ii) $\mathbf{E}[|f_1(X_{i1}, \mathbf{Z}_{i1})||\mathbf{Z}_{i1}]$ is uniformly bounded almost surely.

Assumption FE1 imposes several independence conditions. Assumption FE1(i) requires strict exogeneity, a standard condition in the panel data literature. Assumption FE1(ii) serves as a normalization, and Assumption FE1(iii) mirrors Assumption RE1(iv). Notably, Assumption FE1 implies that X_{i1} and (η_{i1}, η_{i2}) are independent conditional on $\mathbf{Z}_{i1}$, a property essential for applying conditional deconvolution.

Assumption FE2 parallels Assumption RE2 and allows the use of fully observed data to recover the conditional distribution of $f_1(X_{i1}, \mathbf{Z}_{i1})$. Assumptions FE3(i) and (ii) coincide with Assumption RE3, while Assumption FE3(iii) provides a location normalization that enables the unique separation of α_i from $f_t(\cdot)$ (Porter, 1996; Henderson, Carroll, and Li, 2008). Assumption FE4 is the fixed effects analog of Assumption RE4 and is standard.

THEOREM 2. *If Assumptions FE1–FE4 are satisfied, then $f_t(x_t, \mathbf{z}_t)$ is identified for all $(x_t, \mathbf{z}_t)$ on the support of $(X_{it}, \mathbf{Z}_{it})$, for $t \in \{1, 2, 3\}$. Moreover, the distribution of X_{i1} is identified.*

The proof of Theorem 2 follows logic similar to that of Theorem 1. We use the completely observed data from the second and third periods to "impute" the conditional distribution of $f_1(X_{i1}, \mathbf{Z}_{i1})$, which enables the identification of $f_1(x_1, \mathbf{z}_1)$ and the distribution of X_{i1}. The formal proof is provided in Appendix B.

Notably, the identification of the random effects model requires only one period of completely observed data, whereas the fixed effects model necessitates two. This

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

difference arises from the need to recover the relevant error distributions used in conditional deconvolution. In the random effects model, where α_i is independent of the regressors, one period suffices. In the fixed effects model, however, dependence between α_i and the regressors must be addressed. We employ first-differencing to eliminate α_i and recover the relevant distributions, a procedure that requires two fully observed periods. This distinction is consistent with standard panel data analysis (Wooldridge, 2010, Chap. 10).

3. ESTIMATION

This section presents the estimation results for the random and fixed effects models. We first construct easy-to-implement estimators following the identification strategies in Section 2 and then derive their uniform convergence rates.

3.1. Random Effects Model

3.1.1. *Estimator.* First, we use the NW kernel estimator to estimate $f_2(x_2, \mathbf{z}_2)$ as follows[2]:

$$\hat{f}_2(x_2, \mathbf{z}_2) = \hat{\mathbf{E}}[Y_{i2}|X_{i2} = x_2, \mathbf{Z}_{i2} = \mathbf{z}_2] = \frac{\sum_{i=1}^n Y_{i2} K_{i,h_{fr}}(x_2, \mathbf{z}_2)}{\sum_{i=1}^n K_{i,h_{fr}}(x_2, \mathbf{z}_2)}, \tag{2}$$

where $K_{i,h_{fr}}(x_2, \mathbf{z_2}) = \frac{1}{h_{fr}^{p+1}} \times k(\frac{X_{i2}-x_2}{h_{fr}}) \times k(\frac{Z_{i2,1}-z_{2,1}}{h_{fr}}) \times \cdots \times k(\frac{Z_{i2,p}-z_{2,p}}{h_{fr}})$. In this expression, $k(\cdot)$ denotes a univariate kernel function, and h_{fr} is the bandwidth (Li and Racine, 2007, Chap. 2). While variable-specific bandwidths can be used, we adopt a uniform bandwidth to simplify the notation for convergence rates.

Next, following Equation (B.2), we estimate the conditional CDF of $f_1(X_{i1}, \mathbf{Z}_{i1})$ as follows:

$$\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1, \mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1)$$
$$= \frac{1}{2} - \int_{-\infty}^{\infty} \frac{\exp(-isf_1(x_1, \mathbf{z}_1))}{2\pi is} \phi_r(h_r s) \frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)} ds, \tag{3}$$

where $\phi_r(\cdot)$ is the Fourier transform of a kernel function $K_r(\cdot)$ with bounded support, and h_r is a bandwidth parameter converging to zero (Fan, 1991b; Evdokimov, 2010). Because deconvolution is an ill-posed problem, $\phi_r(\cdot)$ serves as a smoothing function. We estimate $\phi_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)$ in Equation (3) as follows:

$$\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \frac{\sum_{i=1}^n e^{isY_{i1}} K_{i,h_Y}(\mathbf{z}_1)}{\sum_{i=1}^n K_{i,h_Y}(\mathbf{z}_1)},$$

[2]Other kernel estimators, such as local linear (LL) or local polynomial, can also be applied. Our simulations include the LL estimator as a comparison to the NW results.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

where $K_{i,h_Y}(\mathbf{z}_1) = \frac{1}{h_Y^p} k(\frac{Z_{i1,1}-z_{1,1}}{h_Y}) \times \cdots \times k(\frac{Z_{i1,p}-z_{1,p}}{h_Y})$ with bandwidth h_Y. While bandwidths may vary across regressors, we maintain a uniform h_Y for simplicity. By identification, we have $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \phi_{\epsilon_{i1}}(s) = \phi_{\epsilon_{i2}}(s) = \phi_{\epsilon_{i2}}(s|X_{i2} = x_2, \mathbf{Z}_{i2} = \mathbf{z}_2)$. This suggests that the characteristic function in the first period can be estimated using the conditional characteristic function in the second period. Therefore, we fix a pair of $(x_2, \mathbf{z}_2)$ and estimate $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)$ (equivalently, $\phi_{\epsilon_{i1}}(s)$) in Equation (3) as follows:

$$\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \hat{\phi}_{\epsilon_{i1}}(s) = \frac{\sum_{i=1}^n e^{is(Y_{i2} - \hat{f}_2(X_{i2}, \mathbf{Z}_{i2}))} K_{i,h_\epsilon}(x_2, \mathbf{z}_2)}{\sum_{i=1}^n K_{i,h_\epsilon}(x_2, \mathbf{z}_2)},$$

where $K_{i,h_\epsilon}(x_2, \mathbf{z}_2) = \frac{1}{h_\epsilon^{p+1}} \times k(\frac{X_{i2}-x_2}{h_\epsilon}) \times k(\frac{Z_{i2,1}-z_{2,1}}{h_\epsilon}) \times \cdots \times k(\frac{Z_{i2,p}-z_{2,p}}{h_\epsilon})$ with bandwidth h_ϵ.[3] Since the estimated conditional CDF of $f_1(X_{i1}, \mathbf{Z}_{i1})$ may not be monotone, we apply the rearrangement technique of Chernozhukov, Fernández-Val, and Galichon (2010) to ensure a strictly increasing function, which improves the accuracy of distributional and quantile estimates, particularly in small samples.

Then, we estimate $f_1(x_1, \mathbf{z}_1)$ as follows:

$$\begin{aligned} \hat{f}_1(x_1, \mathbf{z}_1) &= \hat{F}^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}\Big(\big(\hat{F}_{f_1(X_{i1}, \mathbf{Z}_{i1})}(f_1(x_1, \bar{\mathbf{z}}_1)|\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)\big) \big| \mathbf{Z}_{i1} = \mathbf{z}_1 \Big) \\ &= \hat{F}^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}\Big(\big(\hat{F}_{f_1(X_{i1}, \mathbf{Z}_{i1})}(x_1|\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)\big) \big| \mathbf{Z}_{i1} = \mathbf{z}_1 \Big), \end{aligned} \tag{4}$$

where the conditional CDF $F_{f_1(X_{i1}, \mathbf{Z}_{i1})}(\cdot|\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)$ is estimated via Equation (3), and the conditional quantile $F^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}(\cdot|\mathbf{Z}_{i1} = \mathbf{z}_1)$ is obtained by inverting the corresponding estimated conditional CDF. In practice, $(x_1, \mathbf{z}_1)$ should take values in the interior points to avoid boundary estimation issues.

In many applications, researchers are interested in the unknown functions and their functionals (Matzkin, 2003; Imbens and Newey, 2009). A prominent example is the average marginal effect, denoted by $\mathbf{E}\left[\frac{\partial f_t(X_{it}, \mathbf{Z}_{it})}{\partial x}\right]$, which captures the impact of the primary regressor on the outcome. In our empirical exercise, we examine how elementary students' MV affects computation. To estimate the average marginal effect, we first generate a sample $\{(x_{it}, \mathbf{z}_{it})\}_{i=1}^m$ from the distribution of $(X_{it}, \mathbf{Z}_{it})$. We then evaluate the estimated function and approximate the expectation using a sample average:

$$\hat{\mathbf{E}}\Big[\frac{\partial f_t(X_{it}, \mathbf{Z}_{it})}{\partial x}\Big] = \frac{1}{m}\sum_{i=1}^m \frac{\partial \hat{f}_t(x_{it}, \mathbf{z}_{it})}{\partial x}, \qquad t = 1, 2.$$

Estimators for other functionals can be constructed analogously.

[3] Alternatively, $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)$ can be estimated by averaging the conditional estimator over the distribution of $(X_{i2}, \mathbf{Z}_{i2})$. It can also be estimated by using the unconditional empirical characteristic function $\frac{1}{n}\sum_{i=1}^n e^{is(Y_{i2} - \hat{f}_2(X_{i2}, \mathbf{Z}_{i2}))}$. In this article, we follow the framework of Evdokimov (2010) and use the conditional estimator.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

3.1.2. *Convergence Rate.* The convergence rate (and other asymptotic properties) of $\hat{f}_2(x_2, \mathbf{z}_2)$ is well established and therefore omitted (Li and Racine, 2007, Chap. 2). Following Evdokimov (2010), let $\bar{\mathcal{Z}}_1$ be any set satisfying $\bar{\mathcal{Z}}_1 + B^p_{\epsilon_{\mathcal{Z}_1}}(0) \subset \mathcal{Z}_1$, where $\mathcal{Z}_1$ is the support of $\mathbf{Z}_{i1}$ and $B^p_{\epsilon_{\mathcal{Z}_1}}(0)$ is a p-dimensional ball centered at 0^p with radius $\epsilon_{\mathcal{Z}_1} > 0$. Denote by $Q_{X_{i1}}(\cdot)$ the quantile function of X_{i1}. Under Assumptions CRE1–CRE7 in Appendix A.1, and defining $\chi(h_r^{-1}) = \max_t \sup_{s\in[-h_r^{-1}, h_r^{-1}]} \frac{1}{|\phi_{\epsilon_{it}}(s)|}$, we establish the convergence rate of $\hat{f}_1(x_1, \mathbf{z}_1)$. As is standard in the (conditional) deconvolution literature (Fan, 1991b; Evdokimov, 2010), we defer the derivation of asymptotic normality to future work.[4]

THEOREM 3. *Suppose Assumptions RE1–RE4 and CRE1–CRE7 are satisfied. Further, for some small $\delta > 0$, assume that $f_{X_{i1}}(x_1)$ and $\frac{\partial f_1(x_1,\mathbf{z}_1)}{\partial x_1}$ are bounded away from zero and infinity for all $x_1 \in [Q_{X_{i1}}(\delta/2), Q_{X_{i1}}(1-\delta/2)]$ and $\mathbf{z}_1 \in \mathcal{Z}_1$. Then,*

$$\begin{aligned}
&\sup_{(x_1,\mathbf{z}_1)\in[Q_{X_{i1}}(\delta),Q_{X_{i1}}(1-\delta)]\times\bar{\mathcal{Z}}_1} \left|\hat{f}_1(x_1,\mathbf{z}_1) - f_1(x_1,\mathbf{z}_1)\right| \\
&= O_p\Big(h_r^{k_F} + \big([log(n)/nh_Y^p]^{1/2} + h_r^{-(\bar{k}-1)}h_Y^{\bar{k}}\big)h_r^{-1}\chi(h_r^{-1}) + \max\{1, h_r^{k_F-1}\chi^2(h_r^{-1})\} \\
&\quad \times \big([log(n)/nh_\epsilon^{p+1}]^{1/2} + h_r^{-(\bar{k}-1)}h_\epsilon^{\bar{k}} + [log(n)/nh_{fr}^{p+1}]^{1/2} + (p+1)h_{fr}^{\bar{k}}\big)h_r^{-1}\Big).
\end{aligned}$$

We first derive the convergence rate of $\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1)$. Using Equation (3), we have

$$\begin{aligned}
&\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1) - F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1) \\
&= \Big(\frac{1}{2} - \int_{-1/h_r}^{1/h_r} \frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_r(sh_r)\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}ds \\
&\quad - F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1)\Big) \\
&\quad - \Big(\int_{-1/h_r}^{1/h_r} \frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_r(sh_r)\Big[\frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)} - \frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}\Big]ds\Big).
\end{aligned}$$

[4]Deconvolution is a challenging ill-posed inverse problem, and conducting inference in this framework is widely recognized as difficult (Bissantz et al., 2007; Hall and Horowitz, 2013; Chen and Christensen, 2018). While inference results exist for known error distributions (Fan, 1991a; Bissantz and Holzmann, 2008), many economic applications—including the one studied here—involve unknown error distributions that complicate the problem. In such settings, existing approaches typically rely on specific structures, such as symmetric errors or repeated measurements (Kato and Sasaki, 2018; Kato, Sasaki, and Ura, 2021). By contrast, this article considers a more general nonparametric panel data model without these restrictions. Moreover, conditional deconvolution further increases technical complexity. Given these challenges, we leave inference to future research. By developing conditional deconvolution kernel estimators and establishing their convergence rates, this article provides a rigorous foundation for subsequent investigations.

The author thanks the Editor, Peter C. B. Phillips, for noting that the situation is entirely different with nonstationary data, where consistent estimation and asymptotic theory can be developed using standard methods without concern for ill-posed inversions, due to the strength of the signal in the regressors. For example, Wang and Phillips (2009) show that in functional cointegrating regressions where the regressor is an integrated or near-integrated time series, inverse and ill-posed inverse problems do not arise, allowing the use of simple nonparametric kernel estimators.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

The first term converges at rate $h_r^{k_F}$, corresponding to the regularization bias (Fan, 1991b; Evdokimov, 2010). The second term captures the estimation error associated with estimating the characteristic functions $\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_{i1})$ and $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_{i1})$, and converges at the remaining rate. Since estimating $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ requires the prior estimation of $\mathbf{E}[Y_{i2}|X_{i2},\mathbf{Z}_{i2}]$ (Equation (3)), this rate incorporates the standard uniform convergence rate for nonparametric regression (see, e.g., Li and Racine, 2007, Thm. 2.6). Collectively, these components enable the derivation of the convergence rate for $\hat{f}_1(x_1,\mathbf{z}_1)$. The full proof is provided in Appendix B.

The following example illustrates Theorem 3. Suppose $p=1$, $k=2$, $k_F=2$, and $\gamma=6$. Let ϵ_{i1} and ϵ_{i2} follow $Laplace(0,1)$ distributions (i.e., ordinary-smooth), yielding $\chi(h_r^{-1})=1+(h_r^{-1})^2$. Set the bandwidths as $h_r=\left(\frac{log(n)}{n}\right)^{\frac{1}{24}}$, $h_Y=\left(\frac{log(n)}{n}\right)^{\frac{1}{3}}$, $h_\epsilon=\left(\frac{log(n)}{n}\right)^{\frac{1}{4}}$, and $h_{fr}=\left(\frac{log(n)}{n}\right)^{\frac{2}{7}}$. It is straightforward to verify that all conditions required for Theorem 3 are satisfied. Under these parameters, the convergence rate is $O_p\left(\left(\frac{log(n)}{n}\right)^{\frac{1}{21}}\right)$.

3.2. Fixed Effects Model

3.2.1. *Estimator.* First, we use the NW kernel estimator to estimate $f_2(x_2,\mathbf{z}_2)$ as follows:

$$\begin{aligned}\hat{f}_2(x_2,\mathbf{z}_2)&=\hat{\mathbf{E}}[Y_{i2}-Y_{i3}|X_{i2}=x_2,\mathbf{Z}_{i2}=\mathbf{z}_2,X_{i3}=\bar{x}_3,\mathbf{Z}_{i3}=\bar{\mathbf{z}}_3]\\&=\frac{\sum_{i=1}^n(Y_{i2}-Y_{i3})K_{i,h_{fd}}(x_2,\mathbf{z}_2,\bar{x}_3,\bar{\mathbf{z}}_3)}{\sum_{i=1}^n K_{i,h_{fd}}(x_2,\mathbf{z}_2,\bar{x}_3,\bar{\mathbf{z}}_3)},\end{aligned}\tag{5}$$

where $K_{i,h_{fd}}(x_2,\mathbf{z}_2,\bar{x}_3,\bar{\mathbf{z}}_3)=\frac{1}{h_{fd}^{2p+2}}\times k(\frac{X_{i2}-x_2}{h_{fd}})\times k(\frac{Z_{i2,1}-z_{2,1}}{h_{fd}})\times\cdots\times k(\frac{Z_{i2,p}-z_{2,p}}{h_{fd}})\times k(\frac{X_{i3}-\bar{x}_3}{h_{fd}})\times k(\frac{Z_{i3,1}-\bar{z}_{3,1}}{h_{fd}})\times\cdots\times k(\frac{Z_{i3,p}-\bar{z}_{3,p}}{h_{fd}})$. In this expression, $k(\cdot)$ denotes a univariate kernel function, and h_{fd} is the bandwidth (Li and Racine, 2007, Chap. 2). We estimate $f_3(x_3,\mathbf{z}_3)$ using an analogous NW construction.

Next, following Equation (B.5), we estimate the conditional CDF of $f_1(X_{i1},\mathbf{Z}_{i1})$ as follows:

$$\begin{aligned}&\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\\&=\frac{1}{2}-\int_{-\infty}^{\infty}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_f(h_fs)\frac{\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}ds,\end{aligned}\tag{6}$$

where $\phi_f(\cdot)$ is the Fourier transform of a kernel function $K_f(\cdot)$ with bounded support, and h_f is a bandwidth parameter converging to zero (Fan, 1991b; Evdokimov, 2010). We estimate $\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ in Equation (6) as follows:

$$\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)=\frac{\sum_{i=1}^n e^{is(Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2}))}K_{i,h_{Y_d}}(\mathbf{z}_1)}{\sum_{i=1}^n K_{i,h_{Y_d}}(\mathbf{z}_1)},$$

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

where $K_{i,h_{Yd}}(\mathbf{z}_1) = \frac{1}{h_{Yd}^p} k(\frac{Z_{i1,1}-z_{1,1}}{h_{Yd}}) \times \cdots \times k(\frac{Z_{i1,p}-z_{1,p}}{h_{Yd}})$ with bandwidth h_{Yd}, and $\hat{f}_2(X_{i2}, \mathbf{Z}_{i2})$ follows Equation (5). By identification, we have $\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \phi_{\eta_{i1}-\eta_{i2}}(s) = \phi_{\eta_{i3}-\eta_{i2}}(s) = \phi_{\eta_{i3}-\eta_{i2}}(s|X_{i3} = \bar{x}_3, \mathbf{Z}_{i3} = \bar{\mathbf{z}}_3)$. Therefore, $\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)$ (equivalently, $\phi_{\eta_{i1}-\eta_{i2}}(s)$) in Equation (6) can be estimated as follows:

$$\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \hat{\phi}_{\eta_{i1}-\eta_{i2}}(s) = \frac{\sum_{i=1}^n e^{is(Y_{i3}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2}))} K_{i,h_{\epsilon d}}(\bar{x}_3, \bar{\mathbf{z}}_3)}{\sum_{i=1}^n K_{i,h_{\epsilon d}}(\bar{x}_3, \bar{\mathbf{z}}_3)},$$

where $K_{i,h_{\epsilon d}}(\bar{x}_3, \bar{\mathbf{z}}_3) = \frac{1}{h_{\epsilon d}^{p+1}} \times k(\frac{X_{i3}-\bar{x}_3}{h_{\epsilon d}}) \times k(\frac{Z_{i3,1}-\bar{z}_{3,1}}{h_{\epsilon d}}) \times \cdots \times k(\frac{Z_{i3,p}-\bar{z}_{3,p}}{h_{\epsilon d}})$ with bandwidth $h_{\epsilon d}$, and $\hat{f}_2(X_{i2}, \mathbf{Z}_{i2})$ follows Equation (5). The rearrangement technique of Chernozhukov, Fernández-Val, and Galichon (2010) can be applied to ensure that the estimated conditional CDF of $f_1(X_{i1}, \mathbf{Z}_{i1})$ is strictly increasing.

Then, we estimate $f_1(x_1, \mathbf{z}_1)$ as follows:

$$\begin{aligned}\hat{f}_1(x_1, \mathbf{z}_1) &= \hat{F}^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}\Big(\big(\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1, \bar{\mathbf{z}}_1)|\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)\big)\big|\mathbf{Z}_{i1} = \mathbf{z}_1\Big) \\ &= \hat{F}^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}\Big(\big(\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(x_1|\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)\big)\big|\mathbf{Z}_{i1} = \mathbf{z}_1\Big),\end{aligned} \tag{7}$$

where the conditional CDF $F_{f_1(X_{i1},\mathbf{Z}_{i1})}(.|\mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)$ is estimated via Equation (6), and the conditional quantile $F^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(.|\mathbf{Z}_{i1} = \mathbf{z}_1)$ is obtained by inverting the corresponding estimated conditional CDF. As usual, the point $(x_1, \mathbf{z}_1)$ should be taken from the interior. Additionally, we construct estimators for the functionals of $f_t(\cdot)$ using a procedure analogous to that in the random effects model.

3.2.2. *Convergence Rate.* The convergence rates (and other asymptotic properties) of $\hat{f}_2(x_2, \mathbf{z}_2)$ and $\hat{f}_3(x_3, \mathbf{z}_3)$ are well established and therefore omitted. Under Assumptions CFE1–CFE7 in Appendix A.2, and defining $\chi_d(h_f^{-1}) = \max_t \sup_{s\in[-h_f^{-1},h_f^{-1}]} \frac{1}{|\phi_{\eta_{it}-\eta_{i,t+1}}(s)|}$, we establish the convergence rate of $\hat{f}_1(x_1, \mathbf{z}_1)$.

THEOREM 4. *Suppose Assumptions FE1–FE4 and CFE1–CFE7 are satisfied. Further, for some small $\delta > 0$, assume that $f_{X_{i1}}(x_1)$ and $\frac{\partial f_1(x_1,\mathbf{z}_1)}{\partial x_1}$ are bounded away from zero and infinity for all $x_1 \in [Q_{X_{i1}}(\delta/2), Q_{X_{i1}}(1-\delta/2)]$ and $\mathbf{z}_1 \in \mathcal{Z}_1$. Then,*

$$\begin{aligned}&\sup_{(x_1,\mathbf{z}_1)\in[Q_{X_{i1}}(\delta),Q_{X_{i1}}(1-\delta)]\times\bar{\mathcal{Z}}_1} \Big|\hat{f}_1(x_1, \mathbf{z}_1) - f_1(x_1, \mathbf{z}_1)\Big| \\ &= O_p\Big(h_f^{k_F} + \big([log(n)/nh_{Yd}^p]^{1/2} + h_f^{-(\bar{k}-1)}h_{Yd}^{\bar{k}} + [log(n)/nh_{fd}^{2p+2}]^{1/2} \\ &\quad + (2p+2)h_{fd}^{\bar{k}}\big)h_f^{-1}\chi_d(h_f^{-1}) + \max\{1, h_f^{k_F-1}\chi_d^2(h_f^{-1})\} \\ &\quad \times \big([log(n)/nh_{\epsilon d}^{p+1}]^{1/2} + h_f^{-(\bar{k}-1)}h_{\epsilon d}^{\bar{k}} + [log(n)/nh_{fd}^{2p+2}]^{1/2} + (2p+2)h_{fd}^{\bar{k}}\big)h_f^{-1}\Big).\end{aligned}$$

There are two primary differences between Theorems 3 and 4. First, since $f_2(x_2, \mathbf{z}_2)$ is estimated using $\hat{\mathbf{E}}[Y_{i2}|X_{i2}, \mathbf{Z}_{i2}]$ in the random effects model but via

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

$\hat{\mathbf{E}}[Y_{i2} - Y_{i3}|X_{i2}, \mathbf{Z}_{i2}, X_{i3}, \mathbf{Z}_{i3}]$ in the fixed effects model, Theorem 3 includes the term $[log(n)/nh_{fr}^{p+1}]^{1/2} + (p+1)h_{fr}^{\bar{k}}$, whereas Theorem 4 includes $[log(n)/nh_{fd}^{2p+2}]^{1/2} + (2p+2)h_{fd}^{\bar{k}}$. Second, the term $[log(n)/nh_{fd}^{2p+2}]^{1/2} + (2p+2)h_{fd}^{\bar{k}}$ appears in the second part of the expression in Theorem 4 but is absent from Theorem 3. This difference arises because the numerator of the estimator for $\phi_{f_1(X_{i1}, \mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)$ in the fixed effects model requires a first-stage estimation of $f_2(X_{i2}, \mathbf{Z}_{i2})$, whereas that in the random effects model does not. Aside from these two differences, the remaining components of Theorem 4 closely resemble those of Theorem 3. The full proof is provided in Appendix B.

The following example illustrates Theorem 4. Suppose $p = 1$, $\bar{k} = 2$, $k_F = 2$, and $\gamma = 5$. Let η_{i1}, η_{i2}, and η_{i3} be independent and follow $Normal(0, 1/2)$ distributions (i.e., super-smooth), yielding $\chi_d(h_f^{-1}) = \exp((h_f^{-1})^2/2)$. Set the bandwidths as $h_f = 3log(n)^{-\frac{1}{2}}$, $h_{Yd} = \left(\frac{log(n)}{n}\right)^{\frac{1}{3}}$, $h_{\epsilon d} = \left(\frac{log(n)}{n}\right)^{\frac{1}{4}}$, and $h_{fd} = \left(\frac{log(n)}{n}\right)^{\frac{1}{8}}$. It is straightforward to verify that all conditions required for Theorem 4 are satisfied. Under these parameters, the convergence rate is $O_p(log(n)^{-1})$.

4. SIMULATION

In this section, we conduct two simulations to study the small-sample properties of $\hat{f}_1(x_1, \mathbf{z}_1)$. The first considers the random effects model, and the second examines the fixed effects model.

4.1. Simulation I: Random Effects Model

We consider the following specification:

$$Y_{i1} = X_{i1} + (2 + X_{i1})(2Z_{i1} - 1)^3 + \alpha_i + \eta_{i1};$$

$$Y_{i2} = X_{i2} + Z_{i2} + 0.5X_{i2}Z_{i2}^2 + \alpha_i + \eta_{i2}.$$

The functional form for $f_1(X_{i1}, Z_{i1})$ follows Evdokimov (2010), while the specification for $f_2(X_{i2}, Z_{i2})$ is motivated by our empirical exercise.[5] The variables X_{it} and Z_{it} are drawn from $Uniform(0, 1)$ distributions, and α_i follows a standard normal distribution. We assume that η_{i1} and η_{i2} are standard normal (super-smooth). The independence between the persistent heterogeneity α_i and the regressors (X_{it}, Z_{it}) characterizes this as a random effects model.

We run 500 repetitions with sample sizes of 2,500, 5,000, and 10,000 observations. In each repetition, we estimate $f_1(x_1, z_1)$ at 200 fixed points (x_1, z_1) drawn from the 10th to 90th percentiles of their distributions. We use the NW kernel estimator from Section 3 and the LL kernel estimator, the latter of which is less sensitive to boundary bias. Following Evdokimov (2010), we consider two

[5] In the empirical exercise, we examine the effect of MV on computation, controlling for age and SES. Preliminary polynomial regressions, guided by the functional form of $f_1(\cdot)$, indicate that MV, age, and the interaction (with age squared) are significant. Based on these findings, we specify $f_2(\cdot)$ as a polynomial with $(X_{i2}, Z_{i2}, X_{i2}Z_{i2}^2)$ as regressors.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

TABLE 1. Random effects model with X_{i1} completely missing

		$N = 2,500$		$N = 5,000$		$N = 10,000$	
Estimator	BW	*RMSE*	$RBIAS^2$	*RMSE*	$RBIAS^2$	*RMSE*	$RBIAS^2$
NW	0.4	0.2273	0.1307	0.1677	0.1032	0.1554	0.0975
	0.6	0.3492	0.3478	0.3225	0.3187	0.2925	0.2801
LL	0.4	0.1952	0.1192	0.1363	0.0823	0.1345	0.0776
	0.6	0.3076	0.3061	0.2816	0.2752	0.2608	0.2494
Infeasible		0.1029	0.0265	0.0800	0.0224	0.0648	0.0173

bandwidth settings for deconvolution. First is $h_\epsilon = h_{fr} = h_r = 0.4$, followed by $h_\epsilon = h_{fr} = h_r = 0.6$.[6] We also include an "infeasible" benchmark, where $f_1(x_1, z_1)$ is estimated using the NW kernel under full observability of X_{i1}. Table 1 reports the root mean squared error (*RMSE*) and root mean squared bias ($RBIAS^2$), defined as follows:

$$RMSE = \sqrt{\frac{1}{200}\sum_{j=1}^{200}\frac{1}{500}\sum_{r=1}^{500}\left(\hat{f}_{1r}(x_{1j}, z_{1j}) - f_1(x_{1j}, z_{1j})\right)^2},$$

$$RBIAS^2 = \sqrt{\frac{1}{200}\sum_{j=1}^{200}\left(\frac{1}{500}\sum_{r=1}^{500}\hat{f}_{1r}(x_{1j}, z_{1j}) - f_1(x_{1j}, z_{1j})\right)^2}.$$

4.2. Simulation II: Fixed Effects Model

We consider the following specification:

$$Y_{i1} = X_{i1} + (2 + X_{i1})(2Z_{i1} - 1)^3 + \alpha_i + \eta_{i1};$$

$$Y_{i2} = X_{i2} + Z_{i2} + 0.5X_{i2}Z_{i2}^2 + \alpha_i + \eta_{i2};$$

$$Y_{i3} = -1 + X_{i3} + Z_{i3} + (2X_{i3} - 1)^3 Z_{i3} + (2Z_{i3} - 1)^3 X_{i3} + \alpha_i + \eta_{i3}.$$

The functional forms for $f_1(\cdot)$ and $f_2(\cdot)$ are identical to those in the random effects model, and the specification for $f_3(\cdot)$ follows Evdokimov (2010). The variables X_{it} and Z_{it} are drawn from $Uniform(0, 1)$ distributions, and the unobserved persistent heterogeneity is defined as $\alpha_i = 0.5\sum_{t=1}^{3}(X_{it} - 0.5) + 0.5\,Normal(0, 1)$. We assume that η_{i1}, η_{i2}, and η_{i3} follow standard normal distributions. Because the unobserved

[6] While several methods have been proposed for bandwidth selection in deconvolution problems—such as minimizing the MISE (Stefanski and Carroll, 1990) or its asymptotically dominant part (Delaigle and Gijbels, 2004)—established methods for conditional deconvolution remain limited. Because bandwidth selection is not our primary focus, we follow Evdokimov (2010) and report results for 0.4 and 0.6, noting that smaller bandwidths (e.g., 0.2) provide insufficient regularization and lead to poor performance.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

Table 2. Fixed effects model with X_{i1} completely missing

		$N = 2,500$		$N = 5,000$		$N = 10,000$	
Estimator	BW	*RMSE*	$RBIAS^2$	*RMSE*	$RBIAS^2$	*RMSE*	$RBIAS^2$
NW	0.4	0.3317	0.2045	0.2665	0.1498	0.1965	0.1138
	0.6	0.3755	0.3708	0.3571	0.3476	0.3247	0.3192
LL	0.4	0.2847	0.1792	0.2278	0.1131	0.1678	0.0991
	0.6	0.3353	0.3308	0.3146	0.3075	0.2805	0.2747
Infeasible		0.1811	0.0837	0.1591	0.0608	0.1315	0.0520

persistent heterogeneity is correlated with the regressors, this constitutes a fixed effects model.

Consistent with the previous simulation, we run 500 repetitions with sample sizes of 2,500, 5,000, and 10,000 observations. In each repetition, we estimate $f_1(x_1, z_1)$ at 200 fixed points (x_1, z_1) drawn from the 10th to 90th percentiles of their distributions. We use the NW and LL kernel estimators and consider two bandwidth settings for deconvolution: 0.4 and 0.6. We also include an "infeasible" benchmark estimator. Table 2 reports the *RMSE* and $RBIAS^2$, defined previously.

The simulation results show that the estimators perform well in small samples. Several additional patterns emerge. First, a bandwidth of 0.4 outperforms 0.6 in both models, yielding smaller and more balanced *RMSE* and $RBIAS^2$ values, consistent with the conditional deconvolution literature. Second, performance improves with sample size. Third, the NW and LL kernel estimators produce similar results, as the evaluation points exclude boundary regions. Finally, Simulation I exhibits better results than Simulation II because the latter applies first-differencing to eliminate unobserved persistent heterogeneity, introducing additional conditioning variables; this result is consistent with Theorems 3 and 4.

5. EMPIRICAL EXERCISE

Understanding the mechanisms underlying mathematical difficulties is essential for developing competencies required for academic and professional success (Peng and Lin, 2019; Lin, Peng, and Luo, 2021). In this section, we examine the effect of MV on computation (COM) among fourth- and sixth-grade Chinese elementary students. Section 5.1 describes the data, and Section 5.2 presents the estimation results.

5.1. Data

We use a dataset from Lin, Peng, and Luo (2021) comprising 237 participants (125 boys) from a representative elementary school in southern China. The school was selected because, according to county education bureau records, its

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

TABLE 3. Summary statistics of data

	Minimum	25%	Median	Mean	75%	Maximum
COM_4	21.00	35.00	39.00	38.60	42.00	55.00
COM_6	24.00	63.00	68.00	66.45	73.00	79.00
MV_4	15.00	35.62	43.00	42.68	50.00	65.00
Age_4	104.00	119.00	122.00	122.30	125.00	151.00
Age_6	152.00	167.00	170.00	170.30	173.00	199.00
$SES_{4/6}$	0.50	2.00	2.50	2.51	3.50	4.50

academic performance aligns with the county average. MV is measured using multiple-choice questions covering the relevant vocabulary for fourth and sixth grades, where students select either the correct definition of a term or the term corresponding to a given definition. COM is assessed using the Wide Range Achievement Test, in which students have 30 minutes to complete 80 calculation problems of increasing difficulty. For this test, a small number of English terms were translated into Chinese. Scores are defined as the number of correctly solved items.

Computation data are available for both the fourth and sixth grades. However, MV data were collected only for the fourth grade, leaving sixth-grade observations completely missing due to funding and time constraints.[7] The regressors $\mathbf{Z}_{it}$ include age (measured in months) and SES, both of which are fully observed for both grades. Table 3 reports the summary statistics.

5.2. Estimation Results

We employ the following nonparametric random effects panel data model[8]:

$$COM_{it} = f_t(MV_{it}, \mathbf{Z}_{it}) + \epsilon_{it}, \quad t \in \{4, 6\}.$$

Applying the identification and estimation strategies developed in this article, we examine the average marginal effects of MV on COM, defined as $\mathbf{E}[\frac{\partial f_t(MV_{it}, \mathbf{Z}_{it})}{\partial mv}]$ for $t \in \{4, 6\}$. For the sixth-grade sample, we set the deconvolution bandwidths to 0.4 and 0.6 to maintain consistency with the simulations. The first column of Table 4 reports the results. For fourth-grade students, the estimated effect is 0.251 (s.e. 0.124). For sixth-grade students, the estimates exhibit some sensitivity to bandwidth choice: the estimate closest to the fourth-grade result is 0.244

[7] Owing to this missingness, Lin, Peng, and Luo (2021) focus primarily on fourth-grade students; our analysis therefore complements their work.

[8] The random effects specification is used due to the two-period data constraint; however, fixed effects models may be feasible with longer panels.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

Table 4. Effects of MV on COM in different grades

	$\mathbf{E}[\frac{\partial f_t(MV_{it},\mathbf{Z}_{it})}{\partial mv}]$	$\mathbf{E}[\frac{\partial f_t(MV_{it},\mathbf{Z}_{it})}{\partial mv}\|MV_{it}\in\mathcal{Q}_j]$			
		0%–25%	25%–50%	50%–75%	75%–100%
Fourth grade	0.251	0.233	0.239	0.253	0.278
	(0.124)	(0.089)	(0.139)	(0.126)	(0.136)
Sixth grade ($bw = 0.4$)	0.244	0.215	0.223	0.275	0.322
	(0.145)	(0.087)	(0.146)	(0.145)	(0.161)
Sixth grade ($bw = 0.6$)	0.302	0.242	0.321	0.267	0.355
	(0.175)	(0.103)	(0.182)	(0.149)	(0.218)

Note: Standard errors are reported in parentheses.

(s.e. 0.145), obtained with a bandwidth of 0.4.[9] These findings highlight the important role of MV in developing computational proficiency.

Compared with linear models, the nonparametric approach offers the advantage of capturing more complex effects of MV on COM. We partition the support of MV into empirical quartiles, $\mathcal{Q}_j$, defined by the intervals $[F^{-1}_{MV_{it}}(\frac{j-1}{4}), F^{-1}_{MV_{it}}(\frac{j}{4})]$ for $j = 1,\ldots,4$. We then estimate the average marginal effects within each segment: $\mathbf{E}\left[\frac{\partial f_t(MV_{it},\mathbf{Z}_{it})}{\partial mv}|MV_{it}\in\mathcal{Q}_j\right]$. Columns 2–5 of Table 4 show that these effects vary with the level of MV. For both fourth- and sixth-grade students, the marginal effects increase with MV, indicating that MV has a progressively stronger impact on computational performance.

6. CONCLUSION

The nonparametric time-varying panel data model has gained considerable attention in recent years. However, identification and estimation become challenging when primary regressors are completely missing in some time periods due to funding constraints, time limitations, storage capacity, or legal requirements. This article proposes a novel strategy to address these challenges. Under standard assumptions from the panel data and nonparametric literature, we establish the identification of the unknown functions and the distributions of the missing regressors. Specifically, the random effects model is identified with one period of completely observed data, whereas the fixed effects model requires two. Following identification, we construct estimators: in periods with fully observed data, estimation relies on conditional expectations; in periods with missing data,

[9]Because (conditional) deconvolution is an ill-posed problem with limited formal inference methods, we follow the literature and use the nonparametric bootstrap to compute standard errors (Krasnokutskaya, 2011; Kato, Sasaki, and Ura, 2021). Specifically, we draw 500 bootstrap samples by resampling individuals with replacement from the original sample, repeat the entire estimation procedure to obtain the marginal effect for each bootstrap sample, and compute the standard deviation of the bootstrap estimates.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

it combines conditional CDFs and quantiles. We derive uniform convergence rates for these estimators and demonstrate favorable small-sample performance through simulation studies. An empirical application examining the effect of MV on computation further illustrates the relevance and applicability of the proposed method.

APPENDIX

A. ASSUMPTIONS FOR CONVERGENCE RATES

A.1. Random Effects Model Assumptions

We impose smoothness and regularity assumptions on the random variables and the unknown functions (Fan, 1991b; Evdokimov, 2010). A function $\psi : \mathbf{S} \times \mathbf{W} \to \mathbb{C}$, where $\mathbf{S} \subset \mathbb{R}$ and $\mathbf{W} \subset \mathbb{R}^p$, is said to belong to the class of functions $\mathbf{D}^{\mathbf{S}}_{\mathbf{W}}(k)$ if it is k times continuously differentiable in w, and there is a constant $A > 0$ such that $\sup_{w\in\mathbf{W}} \max_{|\alpha|=l} |\partial^{\alpha}_{w}\psi(s,w)| \leq A(1+|s|^{l})$ for all $l \leq k$ and all s on the support. Here, $\partial^{\alpha}_{w}\psi(s,w)$ is defined as $\partial^{|\alpha|}\psi(s,w)/\partial w_1^{\alpha_1}\ldots\partial w_p^{\alpha_p}$ with $|\alpha| = \alpha_1 + \cdots + \alpha_p$.

Assumption CRE1. $F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1)$ has k_F continuous absolutely integrable derivatives for all $f_1(x_1,\mathbf{z}_1)$ and $\mathbf{z}_1$ on the support.

Assumption CRE2. For $t \in \{1,2\}$: (i) the support of $(X_{it},\mathbf{Z}_{it})$ is bounded; (ii) $f_{X_{it},\mathbf{Z}_{it}}(x_t,\mathbf{z}_t)$ is bounded from above and away from zero for all $(x_t,\mathbf{z}_t)$ on the support; and (iii) there is a positive integer $\bar{k} \geq 1$ such that $f_{X_{it},\mathbf{Z}_{it}}(x_t,\mathbf{z}_t)$ has $\bar{k}$ continuous mixed partial derivatives.

Assumption CRE3. For $t \in \{1,2\}$: (i) there is a constant $\gamma > 2$ such that $\sup_{x_t,\mathbf{z}_t} \mathbf{E}[|Y_{it}|^{\gamma}|X_{it} = x_t, \mathbf{Z}_{it} = \mathbf{z}_t]$ is bounded and (ii) $\mathbf{E}[Y_{it}|X_{it} = x_t, \mathbf{Z}_{it} = \mathbf{z}_t]$ is $\bar{k}$-th differentiable.

Assumption CRE4. $\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) \in \mathbf{D}^{[-h_r^{-1},h_r^{-1}]}_{\mathcal{Z}_1}(\bar{k})$, $\partial\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)/\partial s \in \mathbf{D}^{[-h_r^{-1},h_r^{-1}]}_{\mathcal{Z}_1}(\bar{k})$.

Assumption CRE1 is stronger than the conventional assumption of bounded derivatives and is needed to establish the convergence rate of the Fourier regularization bias. Assumptions CRE2 and CRE3 are standard conditions on the support, the smoothness of the density and the regression function, and the conditional moment of the outcome variable for deriving uniform convergence rate. Assumption CRE4 imposes smoothness conditions on $\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)$ and its first derivative. These conditions are required to establish the uniform convergence rate of the weighted characteristic function estimation error appearing in the deconvolution step.

The tail behavior of the characteristic function of ϵ_{it} is critical in determining the difficulty of deconvolution. As discussed in Fan (1991b), a smoother distribution of ϵ_{it} leads to a more challenging deconvolution. Two categories are typically considered: ordinary-smooth (e.g., Gamma and Laplace) and super-smooth (e.g., Normal and Cauchy). In our simulations, we consider that ϵ_{it} follows a normal distribution. Recalling the notation

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

introduced in the main text, we write

$$\chi(h_r^{-1}) = \max_t \sup_{s\in[-h_r^{-1},h_r^{-1}]} \frac{1}{|\phi_{\epsilon_{it}}(s)|}.$$

The next three assumptions are on the kernel functions and bandwidths used in estimation (Fan, 1991b; Evdokimov, 2010).

Assumption CRE5. The Fourier transform $\phi_r(\cdot)$ is continuously differentiable, has support [-1,1], and satisfies $\phi_r(s) = 1 + O(|s|^{k_F})$ as $s \to 0$.

Assumption CRE6. $k(\cdot)$ is a bounded, integrable, univariate $\bar{k}$-th-order kernel function.

Assumption CRE7.

(i) $h_Y \to 0, nh_Y \to \infty, h_\epsilon \to 0, nh_\epsilon \to \infty, h_{fr} \to 0, nh_{fr} \to \infty, h_r \to 0.$
(ii) $(log(n))^{1/2} n^{1/\gamma - 1/2} h_Y^{-p/2} \to 0, (log(n))^{1/2} n^{1/\gamma-1/2} h_\epsilon^{-(p+1)/2} \to 0.$
(iii) $(log(n))^{\gamma-3/2} h_r^{-1} h_Y^{p/2} n^{1/\gamma-1/2} \to 0, (log(n))^{\gamma-3/2} h_r^{-1} h_\epsilon^{(p+1)/2} n^{1/\gamma-1/2} \to 0.$
(iv) $h_r^{-1}\big([log(n)/nh_\epsilon^{p+1}]^{1/2} + h_r^{-(\bar{k}-1)} h_\epsilon^{\bar{k}} + [log(n)/nh_{fr}^{p+1}]^{1/2} + (p+1)h_{fr}^{\bar{k}}\big)\chi^2(h_r^{-1}) \to 0.$

Assumption CRE5 concerns the Fourier regularization function, and Assumption CRE6 specifies the univariate kernel function used in estimation; both are standard conditions. Assumptions CRE7(i) and (iv) are needed for consistency. In particular, Assumption CRE7(iv) ensures that $\sup_{(s,\mathbf{z}_1)\in[-1/h_r,1/h_r]\times\mathcal{Z}_1} \left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1) - \phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right| = o_p(1)$, which is necessary because $\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ appears in the denominator of the estimator $\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$. Assumptions CRE7(ii) and (iii) require the tails of Y_{it} to be sufficiently thin. Although these conditions are not necessary for consistency, they simplify the expression for the uniform convergence rate.

A.2. Fixed Effects Model Assumptions

The following assumptions are needed to derive the convergence rates of the fixed effects model. Recall that $\chi_d(h_f^{-1})$ is defined as $\max_t \sup_{s\in[-h_f^{-1},h_f^{-1}]} \frac{1}{|\phi_{\eta_{it}-\eta_{i,t+1}}(s)|}$.

Assumption CFE1. $F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$ has k_F continuous absolutely integrable derivatives for all $f_1(x_1,\mathbf{z}_1)$ and $\mathbf{z}_1$ on the support.

Assumption CFE2. For $t \in \{1,2,3\}$: (i) the support of $(X_{it},\mathbf{Z}_{it})$ is bounded; (ii) $f_{X_{it},\mathbf{Z}_{it}}(x_t,\mathbf{z}_t)$ is bounded from above and away from zero for all $(x_t,\mathbf{z}_t)$ on the support; and (iii) there is a positive integer $\bar{k} \geq 1$ such that $f_{X_{it},\mathbf{Z}_{it}}(x_t,\mathbf{z}_t)$ has $\bar{k}$ continuous mixed partial derivatives.

Assumption CFE3. (i) There is a constant $\gamma > 2$ such that $\sup_{x_1,x_2,\mathbf{z}_1,\mathbf{z}_2} \mathbf{E}[|Y_{i1} - Y_{i2}|^\gamma | X_{i1}=x_1, X_{i2}=x_2, \mathbf{Z}_{i1}=\mathbf{z}_1, \mathbf{Z}_{i2}=\mathbf{z}_2]$ and $\sup_{x_2,x_3,\mathbf{z}_2,\mathbf{z}_3} \mathbf{E}[|Y_{i3} - Y_{i2}|^\gamma | X_{i2}=x_2, X_{i3}=x_3, \mathbf{Z}_{i2}=\mathbf{z}_2, \mathbf{Z}_{i3}=\mathbf{z}_3]$ are bounded. (ii) $\mathbf{E}[Y_{i3} - Y_{i2} | X_{i2}=x_2, \mathbf{Z}_{i2}=\mathbf{z}_2, X_{i3}=x_3, \mathbf{Z}_{i3}=\mathbf{z}_3]$ is $\bar{k}$-th differentiable.

Assumption CFE4. $\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1) \in \mathbf{D}_{\mathcal{Z}_1}^{[-h_f^{-1},h_f^{-1}]}(\bar{k})$, $\partial\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)/\partial s \in \mathbf{D}_{\mathcal{Z}_1}^{[-h_f^{-1},h_f^{-1}]}(\bar{k})$.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

Assumption CFE5. The Fourier transform $\phi_f(\cdot)$ is continuously differentiable, has support [-1,1], and satisfies $\phi_f(s) = 1 + O(|s|^{k_F})$ as $s \to 0$.

Assumption CFE6. $k(\cdot)$ is a bounded, integrable, univariate $\bar{k}$-th-order kernel function.

Assumption CFE7.

(i) $h_{Yd} \to 0$, $nh_{Yd} \to \infty$, $h_{\epsilon d} \to 0$, $nh_{\epsilon d} \to \infty$, $h_{fd} \to 0$, $nh_{fd} \to \infty$, $h_f \to 0$.
(ii) $(log(n))^{1/2} n^{1/\gamma - 1/2} h_{Yd}^{-p/2} \to 0$, $(log(n))^{1/2} n^{1/\gamma - 1/2} h_{\epsilon d}^{-(p+1)/2} \to 0$.
(iii) $(log(n))^{\gamma - 3/2} h_f^{-1} h_{Yd}^{p/2} n^{1/\gamma - 1/2} \to 0$, $(log(n))^{\gamma - 3/2} h_f^{-1} h_{\epsilon d}^{(p+1)/2} n^{1/\gamma - 1/2} \to 0$.
(iv) $h_f^{-1}\left([log(n)/nh_{\epsilon d}^{p+1}]^{1/2} + h_f^{-(\bar{k}-1)} h_{\epsilon d}^{\bar{k}} + [log(n)/nh_{fd}^{2p+2}]^{1/2} + (2p+2)h_{fd}^{\bar{k}}\right)\chi_d^2(h_f^{-1}) \to 0$.

Assumption CFE2 extends Assumption CRE2 from two to three time periods, as required for the identification of the fixed effects model. Assumption CFE3 generalizes the conditional moment condition in Assumption CRE3 from single-period outcomes to first-differenced outcomes. The bandwidth conditions in Assumption CFE7 are modified to accommodate the additional conditioning variables arising from first differencing. Assumptions CFE1 and CFE4–CFE6 are analogous to those for the random effects model, differing only in notation.

B. PROOFS

B.1. Proof of Theorem 1

Proof. Under Assumption RE1, the conditional expectation $\mathbf{E}[Y_{i2}|X_{i2} = x_2, \mathbf{Z}_{i2} = \mathbf{z}_2]$ identifies $f_2(x_2, \mathbf{z}_2)$. We address the identification of $f_1(x_1, \mathbf{z}_1)$ by first recovering the characteristic function of ϵ_{i1}, denoted by $\phi_{\epsilon_{i1}}(s)$. Under Assumptions RE1, RE2, and RE4, we have

$$\phi_{\epsilon_{i1}}(s) = \phi_{\epsilon_{i2}}(s) = \phi_{Y_{i2} - \mathbf{E}[Y_{i2}|X_{i2}, \mathbf{Z}_{i2}]}(s). \tag{B.1}$$

Next, we identify the characteristic function of $f_1(X_{i1}, \mathbf{Z}_{i1})$ conditional on $\mathbf{Z}_{i1} = \mathbf{z}_1$ for any $\mathbf{z}_1$ on the support. By Assumptions RE1 and RE4, we obtain

$$\phi_{f_1(X_{i1}, \mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1)}.$$

The numerator is identified from the data, and the denominator is identified by Equation (B.1), noting that $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) = \phi_{\epsilon_{i1}}(s)$. This conditional characteristic function identifies the conditional CDF as follows:

$$\begin{aligned} &F_{f_1(X_{i1}, \mathbf{Z}_{i1})}(f_1(x_1, \mathbf{z}_1)|\mathbf{Z}_{i1} = \mathbf{z}_1) \\ &= \frac{1}{2} - \lim_{\chi \to \infty} \int_{-\chi}^{\chi} \frac{\exp(-isf_1(x_1, \mathbf{z}_1))}{2\pi is} \phi_{f_1(X_{i1}, \mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1} = \mathbf{z}_1) ds. \end{aligned} \tag{B.2}$$

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

Finally, we identify $f_1(x_1, \mathbf{z}_1)$ using the identified conditional CDF and Assumptions RE1 and RE3. Specifically, it follows that

$$\begin{aligned}
&Pr(f_1(X_{i1}, \mathbf{Z}_{i1}) \leq f_1(x_1, \mathbf{z}_1) | \mathbf{Z}_{i1} = \mathbf{z}_1) \\
&= Pr(X_{i1} \leq x_1 | \mathbf{Z}_{i1} = \mathbf{z}_1) = Pr(X_{i1} \leq x_1 | \mathbf{Z}_{i1} = \bar{\mathbf{z}}_1) \\
&= Pr(f_1(X_{i1}, \mathbf{Z}_{i1}) \leq x_1 | \mathbf{Z}_{i1} = \bar{\mathbf{z}}_1).
\end{aligned}$$

This implies

$$f_1(x_1, \mathbf{z}_1) = F^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}\Big(\big(F_{f_1(X_{i1}, \mathbf{Z}_{i1})}(x_1 | \mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)\big) \big| \mathbf{Z}_{i1} = \mathbf{z}_1 \Big), \tag{B.3}$$

where $F^{-1}_{(\cdot)}(\cdot|\cdot)$ denotes the conditional quantile function. These results, together with Assumption RE1, establish that the distribution of X_{i1} is also identified. □

B.2. Proof of Theorem 2

Proof. We employ first-differencing to eliminate the unobserved persistent heterogeneity α_i:

$$Y_{i1} - Y_{i2} = f_1(X_{i1}, \mathbf{Z}_{i1}) - f_2(X_{i2}, \mathbf{Z}_{i2}) + \eta_{i1} - \eta_{i2},$$

$$Y_{i3} - Y_{i2} = f_3(X_{i3}, \mathbf{Z}_{i3}) - f_2(X_{i2}, \mathbf{Z}_{i2}) + \eta_{i3} - \eta_{i2}.$$

Under Assumptions FE1–FE3, $\mathbf{E}[Y_{i2} - Y_{i3} | X_{i2} = x_2, \mathbf{Z}_{i2} = \mathbf{z}_2, X_{i3} = \bar{x}_3, \mathbf{Z}_{i3} = \bar{\mathbf{z}}_3]$ identifies $f_2(x_2, \mathbf{z}_2)$. This, in turn, allows $\mathbf{E}[Y_{i3} - Y_{i2} + f_2(X_{i2}, \mathbf{Z}_{i2}) | X_{i3} = x_3, \mathbf{Z}_{i3} = \mathbf{z}_3]$ to identify $f_3(x_3, \mathbf{z}_3)$.

We identify $f_1(x_1, \mathbf{z}_1)$ by first recovering the characteristic function of $\eta_{i1} - \eta_{i2}$, denoted by $\phi_{\eta_{i1} - \eta_{i2}}(s)$. Under Assumptions FE1–FE4, we have

$$\phi_{\eta_{i1} - \eta_{i2}}(s) = \phi_{\eta_{i3} - \eta_{i2}}(s) = \phi_{Y_{i3} - Y_{i2} + f_2(X_{i2}, \mathbf{Z}_{i2})}(s | X_{i3} = \bar{x}_3, \mathbf{Z}_{i3} = \bar{\mathbf{z}}_3). \tag{B.4}$$

Next, we identify the characteristic function of $f_1(X_{i1}, \mathbf{Z}_{i1})$ conditional on $\mathbf{Z}_{i1} = \mathbf{z}_1$ for any $\mathbf{z}_1$ on the support. By Assumptions FE1 and FE4, we obtain

$$\phi_{f_1(X_{i1}, \mathbf{Z}_{i1})}(s | \mathbf{Z}_{i1} = \mathbf{z}_1) = \frac{\phi_{Y_{i1} - Y_{i2} + f_2(X_{i2}, \mathbf{Z}_{i2})}(s | \mathbf{Z}_{i1} = \mathbf{z}_1)}{\phi_{\eta_{i1} - \eta_{i2}}(s | \mathbf{Z}_{i1} = \mathbf{z}_1)}.$$

The numerator is identified from the data and the previous step, and the denominator is identified by Equation (B.4) because $\phi_{\eta_{i1} - \eta_{i2}}(s | \mathbf{Z}_{i1} = \mathbf{z}_1) = \phi_{\eta_{i1} - \eta_{i2}}(s)$. This conditional characteristic function identifies the conditional CDF as follows:

$$\begin{aligned}
&F_{f_1(X_{i1}, \mathbf{Z}_{i1})}(f_1(x_1, \mathbf{z}_1) | \mathbf{Z}_{i1} = \mathbf{z}_1) \\
&= \frac{1}{2} - \lim_{\chi \to \infty} \int_{-\chi}^{\chi} \frac{\exp(-is f_1(x_1, \mathbf{z}_1))}{2\pi i s} \phi_{f_1(X_{i1}, \mathbf{Z}_{i1})}(s | \mathbf{Z}_{i1} = \mathbf{z}_1) ds.
\end{aligned} \tag{B.5}$$

Finally, we identify $f_1(x_1, \mathbf{z}_1)$ using the identified conditional CDF and Assumptions FE1 and FE3:

$$f_1(x_1, \mathbf{z}_1) = F^{-1}_{f_1(X_{i1}, \mathbf{Z}_{i1})}\Big(\big(F_{f_1(X_{i1}, \mathbf{Z}_{i1})}(x_1 | \mathbf{Z}_{i1} = \bar{\mathbf{z}}_1)\big) \big| \mathbf{Z}_{i1} = \mathbf{z}_1 \Big), \tag{B.6}$$

where $F^{-1}_{(\cdot)}(\cdot|\cdot)$ denotes the conditional quantile function. These results, together with Assumption FE1, establish that the distribution of X_{i1} is also identified. □

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

B.3. Proof of Theorem 3

Proof. We first derive the convergence rate of $\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$. By the definition of $\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$, we have

$$\begin{aligned}
&\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)-F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\\
&=\Big(\frac{1}{2}-\int_{-1/h_r}^{1/h_r}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_r(sh_r)\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}ds\\
&\quad-F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\Big)\\
&\quad-\Big(\int_{-1/h_r}^{1/h_r}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_r(sh_r)\Big[\frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}-\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]ds\Big).
\end{aligned}$$

This decomposition consists of two terms. The first term captures the Fourier regularization bias, while the second captures the estimation error arising from replacing $\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ and $\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ with their estimators.

Under Assumptions CRE1 and CRE5, standard arguments from the conditional deconvolution literature, as in Evdokimov (2010), show that the first term satisfies

$$\begin{aligned}
\sup_{(x_1,\mathbf{z}_1)\in\mathcal{X}_1\times\tilde{\mathcal{Z}}_1}\Big|\frac{1}{2}-\int_{-1/h_r}^{1/h_r}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_r(sh_r)\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}ds\\
-F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\Big|=O(h_r^{k_F}).
\end{aligned}$$

We expand $\frac{1}{s}\Big[\frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}-\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]$ appearing in the second term as

$$\begin{aligned}
&\frac{1}{s}\Big[\frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}-\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]\\
&=\Big[\frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\\
&\quad-\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)[\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)]}{s\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)^2}\Big]\\
&\quad\times\Big(1+\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big)^{-1}.
\end{aligned}$$

We define

$$\tilde{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)=\tilde{\phi}_{\epsilon_{i1}}(s)=\frac{\sum_{i=1}^n e^{is(Y_{i2}-f_2(X_{i2},\mathbf{Z}_{i2}))}K_{i,h_\epsilon}(x_2,\mathbf{z}_2)}{\sum_{i=1}^n K_{i,h_\epsilon}(x_2,\mathbf{z}_2)}.$$

Unlike $\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$, which requires estimating $f_2(X_{i2},\mathbf{Z}_{i2})$, $\tilde{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ is the estimator that treats $f_2(X_{i2},\mathbf{Z}_{i2})$ as known.

Under Assumptions CRE1–CRE7, it follows from the uniform convergence rate for nonparametric regression (see, e.g., Li and Racine, 2007, Thm. 2.6) and the convergence results

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

established by Evdokimov (2010) for the conditional characteristic function estimator that

$$\begin{aligned}
&\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_r},\frac{1}{h_r}]\times\bar{\mathcal{Z}}_1}\left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right| \\
&= \sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_r},\frac{1}{h_r}]\times\bar{\mathcal{Z}}_1}\left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\tilde{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)+\tilde{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right| \\
&\le \sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_r},\frac{1}{h_r}]\times\bar{\mathcal{Z}}_1}\left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\tilde{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right| \\
&\quad+\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_r},\frac{1}{h_r}]\times\bar{\mathcal{Z}}_1}\left|\frac{\tilde{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right| \\
&= O_p\Big(h_r^{-1}\big([log(n)/nh_{fr}^{p+1}]^{1/2}+h_{fr}^{\bar{k}}+[log(n)/nh_{\epsilon}^{p+1}]^{1/2}+h_r^{-(\bar{k}-1)}h_{\epsilon}^{\bar{k}}\big)\chi^2(h_r^{-1})\Big)=o_p(1).
\end{aligned}$$

Then, the second term can be bounded as

$$\begin{aligned}
&\sup_{(x_1,\mathbf{z}_1)\in\mathcal{X}_1\times\bar{\mathcal{Z}}_1}\left|\int_{-1/h_r}^{1/h_r}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_r(sh_r)\Big[\frac{\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}-\frac{\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]ds\right| \\
&\le (C+o_p(1))\Big[\Big(\sup_{(s,\mathbf{z}_1)\in[-1/h_r,1/h_r]\times\bar{\mathcal{Z}}_1}\left|s^{-1}\big(\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\big)\right|\Big)h_r^{-1}\chi(h_r^{-1}) \\
&\quad+\Big(\sup_{(s,\mathbf{z}_1)\in[-1/h_r,1/h_r]\times\bar{\mathcal{Z}}_1}|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|\left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right|\Big)h_r^{-1}\Big].
\end{aligned}$$

Under Assumptions CRE1–CRE7, it follows from the convergence results established by Evdokimov (2010) for the conditional characteristic function estimator that

$$\begin{aligned}
&\sup_{(s,\mathbf{z}_1)\in[-1/h_r,1/h_r]\times\bar{\mathcal{Z}}_1}\left|s^{-1}\big(\hat{\phi}_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{Y_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\big)\right| \\
&= O_p([log(n)/nh_Y^p]^{1/2}+h_r^{-(\bar{k}-1)}h_Y^{\bar{k}}).
\end{aligned}$$

Moreover, Assumption CRE1 implies that $|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|<C$ for $|s|\le 1$, and $|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|<C|s|^{-(k_F-1)}$ for $|s|>1$. Combining these decay properties with the preceding convergence result gives

$$\begin{aligned}
&\sup_{(s,\mathbf{z}_1)\in[-1/h_r,1/h_r]\times\bar{\mathcal{Z}}_1}\left(|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|\left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right|\right) \\
&\le O_p\big(([log(n)/nh_{\epsilon}^{p+1}]^{1/2}+h_{\epsilon}^{\bar{k}-1}+[log(n)/nh_{fr}^{p+1}]^{1/2}+(p+1)h_{fr}^{\bar{k}})\big) \\
&\quad+\sup_{\bar{s}\in[1,h_r^{-1}]}\Big[\bar{s}^{1-k_F}\sup_{(s,\mathbf{z}_1)\in[-\bar{s},\bar{s}]\times\bar{\mathcal{Z}}_1}\left|\frac{\hat{\phi}_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\epsilon_{i1}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right|\Big] \\
&= O_p\big(([log(n)/nh_{\epsilon}^{p+1}]^{1/2}+h_{\epsilon}^{\bar{k}-1}+[log(n)/nh_{fr}^{p+1}]^{1/2}+(p+1)h_{fr}^{\bar{k}})\big) \\
&\quad+O_p\Big(\sup_{\bar{s}\in[1,h_r^{-1}]}\big[\bar{s}^{1-k_F}\chi^2(\bar{s})\big([log(n)/nh_{\epsilon}^{p+1}]^{1/2}+\bar{s}^{\bar{k}-1}h_{\epsilon}^{\bar{k}}+[log(n)/nh_{fr}^{p+1}]^{1/2}+h_{fr}^{\bar{k}}\big)\big]\Big)
\end{aligned}$$

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

$$= O_p\big(\max\{1, h_r^{k_F-1}\chi^2(h_r^{-1})\}$$
$$\times ([log(n)/nh_\epsilon^{p+1}]^{1/2} + h_r^{-(\bar{k}-1)}h_\epsilon^{\bar{k}} + [log(n)/nh_{fr}^{p+1}]^{1/2} + (p+1)h_{fr}^{\bar{k}})\big).$$

Putting the bounds for the first and second terms together, the convergence rate of the conditional CDF is

$$\sup_{(x_1,\mathbf{z}_1)\in\mathcal{X}_1\times\bar{\mathcal{Z}}_1} \left|\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1) - F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\right|$$
$$= O_p\Big(h_r^{k_F} + \big([log(n)/nh_Y^p]^{1/2} + h_r^{-(\bar{k}-1)}h_Y^{\bar{k}}\big)h_r^{-1}\chi(h_r^{-1}) + \max\{1, h_r^{k_F-1}\chi^2(h_r^{-1})\}$$
$$\times \big([log(n)/nh_\epsilon^{p+1}]^{1/2} + h_r^{-(\bar{k}-1)}h_\epsilon^{\bar{k}} + [log(n)/nh_{fr}^{p+1}]^{1/2} + (p+1)h_{fr}^{\bar{k}}\big)h_r^{-1}\Big). \tag{B.7}$$

Next, we derive the convergence rate of $\hat{F}^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(q|\mathbf{Z}_{i1}=\mathbf{z}_1)$. Under Assumptions CRE1–CRE7, applying the convergence results for conditional quantile estimators established by Evdokimov (2010), we obtain

$$\sup_{(q,\mathbf{z}_1)\in[\delta,1-\delta]\times\bar{\mathcal{Z}}_1} \left|\hat{F}^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(q|\mathbf{Z}_{i1}=\mathbf{z}_1) - F^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(q|\mathbf{Z}_{i1}=\mathbf{z}_1)\right|$$
$$= O_p\Big(h_r^{k_F} + \big([log(n)/nh_Y^p]^{1/2} + h_r^{-(\bar{k}-1)}h_Y^{\bar{k}}\big)h_r^{-1}\chi(h_r^{-1}) + \max\{1, h_r^{k_F-1}\chi^2(h_r^{-1})\}$$
$$\times \big([log(n)/nh_\epsilon^{p+1}]^{1/2} + h_r^{-(\bar{k}-1)}h_\epsilon^{\bar{k}} + [log(n)/nh_{fr}^{p+1}]^{1/2} + (p+1)h_{fr}^{\bar{k}}\big)h_r^{-1}\Big). \tag{B.8}$$

The conclusion of Theorem 3 follows directly from Equations (B.7) and (B.8). □

B.4. Proof of Theorem 4

Proof. We first derive the convergence rate of $\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$. By the definition of $\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$, we have

$$\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1) - F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)$$
$$= \Big(\frac{1}{2} - \int_{-1/h_f}^{1/h_f} \frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_f(sh_f)\frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}ds$$
$$- F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\Big) - \Big(\int_{-1/h_f}^{1/h_f} \frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_f(sh_f)$$
$$\times \Big[\frac{\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)} - \frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]ds\Big).$$

The first term captures the Fourier regularization bias, while the second captures the estimation error induced by replacing the conditional characteristic functions by their estimators.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

Under Assumptions CFE1 and CFE5, standard arguments from the conditional deconvolution literature, as in Evdokimov (2010), show that the first term satisfies

$$\sup_{(x_1,\mathbf{z}_1)\in\mathcal{X}_1\times\bar{\mathcal{Z}}_1}\Big|\frac{1}{2}-\int_{-1/h_f}^{1/h_f}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_f(sh_f)\frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}ds$$
$$-F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\Big|=O(h_f^{k_F}).$$

We expand $\frac{1}{s}\Big[\frac{\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}-\frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]$ appearing in the second term as

$$\frac{1}{s}\Big[\frac{\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}-\frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]$$
$$=\Big[\frac{\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}$$
$$-\frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)[\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)]}{s\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)^2}\Big]$$
$$\times\Big(1+\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big)^{-1}.$$

We define

$$\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)=\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s)=\frac{\sum_{i=1}^n e^{is(Y_{i3}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2}))}K_{i,h_{\epsilon d}}(\bar{x}_3,\bar{\mathbf{z}}_3)}{\sum_{i=1}^n K_{i,h_{\epsilon d}}(\bar{x}_3,\bar{\mathbf{z}}_3)}.$$

Unlike $\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$, which requires estimating $f_2(X_{i2},\mathbf{Z}_{i2})$, $\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$ is the estimator that treats $f_2(X_{i2},\mathbf{Z}_{i2})$ as known.

Under Assumptions CFE1–CFE7, it follows from the uniform convergence rate for nonparametric regression (see, e.g., Li and Racine, 2007, Thm. 2.6) and the convergence results established by Evdokimov (2010) for the conditional characteristic function estimator that

$$\sup_{(s,\mathbf{z}_1)\in[-1/h_f,1/h_f]\times\bar{\mathcal{Z}}_1}\Big|\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big|$$
$$=\sup_{(s,\mathbf{z}_1)\in[-1/h_f,1/h_f]\times\bar{\mathcal{Z}}_1}$$
$$\Big|\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)+\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big|$$
$$\le\sup_{(s,\mathbf{z}_1)\in[-1/h_f,1/h_f]\times\bar{\mathcal{Z}}_1}\Big|\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big|$$
$$+\sup_{(s,\mathbf{z}_1)\in[-1/h_f,1/h_f]\times\bar{\mathcal{Z}}_1}\Big|\frac{\tilde{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big|$$

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

$$= O_p\Big(h_f^{-1}\big([log(n)/nh_{fd}^{2p+2}]^{1/2} + h_{fd}^{\bar{k}} + [log(n)/nh_{\epsilon d}^{p+1}]^{1/2} + h_f^{-(\bar{k}-1)}h_{\epsilon d}^{\bar{k}}\big)\chi_d^2(h_f^{-1})\Big)$$
$$= o_p(1).$$

Then, the second term can be bounded as

$$\sup_{(x_1,\mathbf{z}_1)\in\mathcal{X}_1\times\bar{\mathcal{Z}}_1}\Big|\int_{-1/h_f}^{1/h_f}\frac{\exp(-isf_1(x_1,\mathbf{z}_1))}{2\pi is}\phi_f(sh_f)$$
$$\Big[\frac{\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)} - \frac{\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big]ds\Big|$$
$$\le \big(C+o_p(1)\big)\Big[\Big(\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}\big|\frac{1}{s}\big(\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$$
$$-\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\big)\big|\Big)h_f^{-1}\chi_d(h_f^{-1})$$
$$+\Big(\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|$$
$$\times\Big|\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\Big|\Big)h_f^{-1}\Big].$$

By the Lipschitz property of complex exponential and L'Hôpital's rule, it holds that

$$\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}$$
$$\Big|\frac{\frac{1}{n}\sum_{i=1}^n\frac{1}{s}\Big(e^{is(Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2}))}K_{i,h_{Y_d}}(\mathbf{z}_1)-e^{is(Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2}))}K_{i,h_{Y_d}}(\mathbf{z}_1)\Big)}{\frac{1}{n}\sum_{i=1}^n K_{i,h_{Y_d}}(\mathbf{z}_1)}\Big|$$
$$\le \sup_{(x_2,\mathbf{z}_2)\in\bar{\mathcal{X}}_2\times\bar{\mathcal{Z}}_2}\Big|\hat{f}_2(x_2,\mathbf{z}_2)-f_2(x_2,\mathbf{z}_2)\Big|\sup_{\mathbf{z}_1\in\bar{\mathcal{Z}}_1}\frac{\frac{1}{n}\sum_{i=1}^n\Big|K_{i,h_{Y_d}}(\mathbf{z}_1)\Big|}{\Big|\frac{1}{n}\sum_{i=1}^n K_{i,h_{Y_d}}(\mathbf{z}_1)\Big|}$$
$$= O_p\big([log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}}\big).$$

This bound, together with the uniform convergence rate of $\hat{\phi}_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)$, yields

$$\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}\Big|s^{-1}\big(\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\big)\Big|$$

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

$$
\begin{aligned}
&\le \sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}\Big|s^{-1}\big(\hat{\phi}_{Y_{i1}-Y_{i2}+\hat{f}_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\\
&\qquad\qquad -\hat{\phi}_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\big)\Big|\\
&+\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}\Big|s^{-1}\big(\hat{\phi}_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\\
&\qquad\qquad -\phi_{Y_{i1}-Y_{i2}+f_2(X_{i2},\mathbf{Z}_{i2})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)\big)\Big|\\
&=O_p\big([log(n)/nh_{Yd}^{p}]^{1/2}+h_f^{-(\bar{k}-1)}h_{Yd}^{\bar{k}}+[log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}}\big).
\end{aligned}
$$

Moreover, Assumption CFE1 implies that $|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|<C$ for $|s|\le 1$, and $|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|<C|s|^{-(k_F-1)}$ for $|s|>1$. Combining these decay properties with the preceding convergence result gives

$$
\begin{aligned}
&\sup_{(s,\mathbf{z}_1)\in[\frac{-1}{h_f},\frac{1}{h_f}]\times\bar{\mathcal{Z}}_1}\left(|\phi_{f_1(X_{i1},\mathbf{Z}_{i1})}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)|\left|\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right|\right)\\
&\le O_p\big([log(n)/nh_{\epsilon d}^{p+1}]^{1/2}+h_{\epsilon d}^{\bar{k}-1}+[log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}}\big)\\
&\quad+\sup_{\bar{s}\in[1,h_f^{-1}]}\left[\bar{s}^{1-k_F}\sup_{(s,\mathbf{z}_1)\in[-\bar{s},\bar{s}]\times\bar{\mathcal{Z}}_1}\left|\frac{\hat{\phi}_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)-\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}{s\phi_{\eta_{i1}-\eta_{i2}}(s|\mathbf{Z}_{i1}=\mathbf{z}_1)}\right|\right]\\
&=O_p\big([log(n)/nh_{\epsilon d}^{p+1}]^{1/2}+h_{\epsilon d}^{\bar{k}-1}+[log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}}\big)\\
&\quad+O_p\Big(\sup_{\bar{s}\in[1,h_f^{-1}]}\big[\bar{s}^{1-k_F}\chi_d^2(\bar{s})\big([log(n)/nh_{\epsilon d}^{p+1}]^{1/2}+\bar{s}^{\bar{k}-1}h_{\epsilon d}^{\bar{k}}\\
&\quad+[log(n)/nh_{fd}^{2p+2}]^{1/2}+h_{fd}^{\bar{k}}\big)\big]\Big)\\
&=O_p\big(\max\{1,h_f^{k_F-1}\chi_d^2(h_f^{-1})\}\\
&\quad\times([log(n)/nh_{\epsilon d}^{p+1}]^{1/2}+h_f^{-(\bar{k}-1)}h_{\epsilon d}^{\bar{k}}+[log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}})\big).
\end{aligned}
$$

Putting the bounds for the first and second terms together, the convergence rate of the conditional CDF is

$$
\begin{aligned}
&\sup_{(x_1,\mathbf{z}_1)\in\mathcal{X}_1\times\bar{\mathcal{Z}}_1}\left|\hat{F}_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)-F_{f_1(X_{i1},\mathbf{Z}_{i1})}(f_1(x_1,\mathbf{z}_1)|\mathbf{Z}_{i1}=\mathbf{z}_1)\right|\\
&=O_p\Big(h_f^{k_F}+\big([log(n)/nh_{Yd}^{p}]^{1/2}+h_f^{-(\bar{k}-1)}h_{Yd}^{\bar{k}}+[log(n)/nh_{fd}^{2p+2}]^{1/2}\\
&\quad+(2p+2)h_{fd}^{\bar{k}}\big)h_f^{-1}\chi_d(h_f^{-1})+\max\{1,h_f^{k_F-1}\chi_d^2(h_f^{-1})\}\\
&\quad\times\big([log(n)/nh_{\epsilon d}^{p+1}]^{1/2}+h_f^{-(\bar{k}-1)}h_{\epsilon d}^{\bar{k}}+[log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}}\big)h_f^{-1}\Big).
\end{aligned}
\tag{B.9}
$$

Next, we derive the convergence rate of $\hat{F}^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(q|\mathbf{Z}_{i1}=\mathbf{z}_1)$. Under Assumptions CFE1–CFE7, applying the convergence results for conditional quantile estimators

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

established by Evdokimov (2010), we obtain

$$\begin{aligned}
&\sup_{(q,\mathbf{z}_1)\in[\delta,1-\delta]\times\bar{\mathcal{Z}}_1}\left|\hat{F}^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(q|\mathbf{Z}_{i1}=\mathbf{z}_1)-F^{-1}_{f_1(X_{i1},\mathbf{Z}_{i1})}(q|\mathbf{Z}_{i1}=\mathbf{z}_1)\right| \\
&= O_p\Big(h_f^{k_F}+\big([log(n)/nh_{Yd}^{p}]^{1/2}+h_f^{-(\bar{k}-1)}h_{Yd}^{\bar{k}}+[log(n)/nh_{fd}^{2p+2}]^{1/2} \\
&\quad+(2p+2)h_{fd}^{\bar{k}}\big)h_f^{-1}\chi_d(h_f^{-1})+\max\{1,h_f^{k_F-1}\chi_d^2(h_f^{-1})\} \\
&\quad\times\big([log(n)/nh_{\epsilon d}^{p+1}]^{1/2}+h_f^{-(\bar{k}-1)}h_{\epsilon d}^{\bar{k}}+[log(n)/nh_{fd}^{2p+2}]^{1/2}+(2p+2)h_{fd}^{\bar{k}}\big)h_f^{-1}\Big).
\end{aligned} \tag{B.10}$$

The conclusion of Theorem 4 follows directly from Equations (B.9) and (B.10). □

FUNDING STATEMENT

The author declares that no specific funding has been received for this article.

COMPETING INTERESTS

The author declares that no competing interests exist.

REFERENCES

Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics*, 93(2), 203–228.

Abrevaya, J. (2019). Missing dependent variables in fixed-effects models. *Journal of Econometrics*, 211(1), 151–165.

Abrevaya, J., & Donald, S. G. (2017). A GMM approach for dealing with missing data on regressors. *Review of Economics and Statistics*, 99(4), 657–662.

Altonji, J. G., & Matzkin, R. L. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica*, 73(4), 1053–1102.

Bissantz, N., Hohage, T., Munk, A., & Ruymgaart, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM Journal on Numerical Analysis*, 45(6), 2610–2636.

Bissantz, N., & Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Problems*, 24(3), 34009.

Chaudhuri, S., & Guilkey, D. K. (2016). GMM with multiple missing variables. *Journal of Applied Econometrics*, 31(4), 678–706.

Chen, B., Yi, G. Y., & Cook, R. J. (2010). Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. *Journal of the American Statistical Association*, 105(489), 336–353.

Chen, X., & Christensen, T. M. (2018). Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric IV regression. *Quantitative Economics*, 9(1), 39–84.

Chernozhukov, V., Fernández-Val, I., & Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3), 1093–1125.

Chernozhukov, V., Fernández-Val, I., Hahn, J., & Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2), 535–580.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica*, 71(5), 1405–1441.

Coe, J. E. (2019). *Estimation of panel data models with missing covariate values* [Ph.D. dissertation]. The University of Texas at Austin.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press

Delaigle, A., & Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics and Data Analysis*, 45(2), 249–267.

Evdokimov, K. (2010). *Identification and estimation of a nonparametric panel data model with unobserved heterogeneity*. Preprint. https://www.kirillevdokimov.com/nppanel_aug12.pdf

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, 53(1), 97–110.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, 19(3), 1257–1272.

Hall, P., & Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Annals of Statistics*, 41(4), 1892–1921.

Henderson, D. J., Carroll, R. J., & Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1), 257–275.

Hu, Y., & Sasaki, Y. (2014). *Recovering the true distribution with paired self reports*. Preprint. http://www.econ2.jhu.edu/people/hu/hu_sasaki_education_2014_03_20.pdf

Hu, Y., & Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1), 195–216.

Imbens, G. W., & Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5), 1481–1512.

Kato, K., & Sasaki, Y. (2018). Uniform confidence bands in deconvolution with unknown error distribution. *Journal of Econometrics*, 207(1), 129–161.

Kato, K., Sasaki, Y., & Ura, T. (2021). Robust inference in deconvolution. *Quantitative*. *Economics*, 12(1), 109–142.

Krasnokutskaya, E. (2011). Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies*, 78(1), 293–327.

Lewbel, A. (2014). An overview of the special regressor method. In J. S. Racine, L. Su, and A. Ullah (Eds.), *The Oxford handbook of applied nonparametric and semiparametric econometrics and statistics* (pp. 38–61). Oxford University Press.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Lin, X., Peng, P., & Luo, H. (2021). The deficit profile of elementary students with computational difficulties versus word problem-solving difficulties. *Learning Disability Quarterly*, 44(2), 110–122.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica*, 71(5), 1339–1375.

Moffit, R., Fitzgerald, J., & Gottschalk, P. (1999). Sample attrition in panel data: The role of selection on observables. *Annales d'Economie et de Statistique*, 55–56, 129–152.

Muris, C. (2020). Efficient GMM estimation with incomplete data. *Review of Economics and Statistics*, 102(3), 518–530.

Peng, P., & Lin, X. (2019). The relation between mathematics vocabulary and mathematics performance among fourth graders. *Learning and Individual Differences*, 69, 11–21.

Porter, J. R. (1996). *Essays in econometrics* [Ph.D. dissertation]. Massachusetts Institute of Technology.

Schennach, S. M. (2013). Measurement error in nonlinear models—A review. *Advances in Economics and Econometrics*, 3, 296–337.

Stefanski, L. A., & Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics*, 21(2), 169–184.

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90(1), 43–52.

Wang, Q., & Phillips, P. C. (2009). Structural nonparametric cointegrating regression. *Econometrica*, 77(6), 1901–1948.

Wilhelm, D. (2015). *Identification and estimation of nonparametric panel data regressions with measurement error*. Preprint. https://www.econstor.eu/bitstream/10419/130050/1/828752974.pdf

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

https://doi.org/10.1017/S0266466626100589 Published online by Cambridge University Press