Errors-In-Variables in Large Nonlinear Panel and Network Models

Kirill S. Evdokimov* Andrei Zeleneev†‡

This version: January 2020

Abstract

We consider estimation of general nonlinear semiparametric panel data models with fixed effects. Estimation of such models implicitly relies on the "within' variation of covariates, which aggravates the Errors-In-Variables (EIV) bias problem. First, we derive the formulas for the bias of m-estimators in large panel data. We show that the bias of common parameters includes both the direct effect of EIV and the EIV bias of the incidental parameters (fixed effects). Then, we propose an estimator that removes the EIV bias in nonlinear models using panel instrumental variables. We show how lagged values of covariates can serve as such instruments in panel data. The estimator does not involve any nonparametric estimation, and can accommodate serially correlated and/or multivariate measurement errors. We establish the asymptotic properties of the estimator. Combined with a jackknife procedure, the estimator is asymptotically normal and unbiased. The properties of the estimator are illustrated in a Monte Carlo simulation. In addition, the estimation approach can be adapted for estimation of large network data models with measurement errors. In particular, we show how the network structure provides instruments needed to eliminate the EIV bias.

*Universitat Pompeu Fabra and Barcelona GSE: kirill.evdokimov@upf.edu.

†Princeton University: zeleneev@princeton.edu.

‡We thank the participants of various seminars and conferences for helpful comments and suggestions. We are also grateful to the Gregory C. Chow Econometrics Research Program at Princeton University and the Department of Economics at the Massachusetts Institute of Technology for their hospitality and support. Evdokimov also acknowledges the financial support from the National Science Foundation via grant SES-1459993, and from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0563).

1 Introduction

The use of panel data in Economics is widespread, thanks to its ability to account for unobserved time-invariant individual heterogeneity. The concern about the potential measurement errors in the data is also widespread. Estimation of the fixed effects models relies on within variation of covariates, which exacerbates the errors-in-variables problem.

The issue of measurement errors in linear panel models is relatively well understood. In the seminal paper, Griliches and Hausman (1986) suggest using lagged values of covariates as readily available instruments. They point out that such instruments are only valid if one is able to difference out the fixed-effects, and investigate the restrictions on the dynamics of the measurement errors that ensures that the model parameters can be identified and estimated.

This paper considers estimation of general nonlinear semiparametric panel data models, such as, e.g., static and dynamic panel probit. Several features of these models make dealing with the measurement errors a hard problem. Addressing the errors-in-variables problem in nonlinear models is known to be difficult by itself. In the nonlinear panel data settings it is further complicated by the incidental parameter problem due to inability to difference out fixed effects. Potential temporal dependence of measurement errors is yet another difficulty.

We propose estimators that are robust to measurement errors. The estimators are easy to compute, and do not require nonparametric estimation or simulation. The estimators have zero mean asymptotically normal distribution, and remove both the measurement error and the incidental parameters biases.

Our results build on a non-standard asymptotic approximation. Let N denote the number of cross-section units, T denote number of time-periods, and Σ denote the variance-covariance matrix of the vector of the measurement errors. We assume that $N \to \infty,\ T \to \infty,\ \|\Sigma\| \to 0$ jointly. The choice of asymptotic approximation determines the scope of applicability of our results. First, we model both N and T as increasing. This is a necessary condition for consistency in general nonlinear panel models.[1] The existing literature suggests that this assumption can provide a reasonably good approximation to finite sample behavior of the estimators as long as $T \gtrapprox 10$ and NT is large. Second, we model variance of measurement errors as shrinking with the sample size. This assumption allows approximate the settings in which bias and standard errors of

[1] Apart from certain particular cases, if, for example, T is treated as fixed, one can only hope to obtain bounds on the parameter values, see Honoré and Tamer (2006) and Chernozhukov et al. (2013).

the estimators are of comparable magnitudes, while allowing sample size to grow.[2] This provides a better approximation of the finite sample properties of estimators and at the same time leads to major simplifications for the estimation procedures. In particular, our approach avoids estimation of any infinite-dimensional nuisance parameters.

We begin our analysis with deriving the bias expressions for the panel data m-estimators in the presence of errors-in-variables. In general, measurement errors not only directly bias the parameters of interest, but also bias the incidental parameters. Since in nonlinear models fixed effects cannot be differenced out, errors-in-variables bias in the incidental parameters generally contributes to the bias of the parameters of interest. Thus, a successful estimation approach needs to address the measurement error bias in the incidental parameters.

One important observation is that in many settings of interest, the attenuation bias and the incidental parameter bias have the opposite signs. In such cases, correcting for incidental parameter bias without correcting the measurement error bias will usually worsen the performance of the estimators. We illustrate this point below in theoretical examples and Monte Carlo experiments.

To leverage the insights of GH86 we set the problem in the more general context of the moment conditions framework, which allows making use of the instrumental variables. Our estimators are panel GMM estimators. These estimators generalize z-estimators for panel data, and, accordingly, suffer from the incidental parameter bias of order T^{-1}. The properties of general large-T panel GMM models have been studied by FL13, and our large sample theory builds on their results.

Measurement error literature. [TBA]

We focus on the large-T settings. [Literature review of large-T panel data here.]

Deriving the bias of the m-estimators requires analyzing the higher-order expansions of the estimators of the incidental parameters as in Rilstone et al. (1996) and Bao and Ullah (2007). We extended their results to the settings with measurement errors.

We note that the analysis using ”small” measurement error is well known. Including Wolter and Fuller (1982); Chesher (1991). Crucially, these papers and a large number of papers in the Statistics literature had either assumed that (relevant features of) the distribution of the measurement error to be known (or perhaps estimable from a separate dataset), or performed sensitivity checks, showing how the estimates vary as functions of

[2]The textbook assumption viewing variance of the measurement errors suggests that in large samples bias is far bigger than standard error. To reconcile smaller magnitudes of the bias one needs to either make assumptions about the values of true parameter vector, or assume that the sample size is not that large.

σ^2. The goal of this paper is to address the typical interest of practitioners that do not have additional information about the distribution of the measurement error, but want to obtain point estimates and valid confidence intervals that account for the EIV bias. The earlier papers did not consider estimation of σ^2 and the related issues of "moderate" measurement error issues that we discuss below. Finally, in contrast to the rest of the literature we consider the settings with incidental parameter problems.

Analysis of Evdokimov and Zeleneev (2016) is appropriate for cross-section and short panel data that can be estimated from the data containing a very small number of time periods. In this paper we address the problem of incidental parameters, and its interaction with problem of measurement error. We derive higher order properties of m-estimators in the presence of measurement error. To the best of our knowledge, this is the first paper to propose feasible estimators for large-T fixed-effect panel data models that are robust to the presence of measurement errors.

Modeling techniques/devices such as local-to-zero approximations (e.g., Staiger and Stock (1997)) have proven themselves to be very useful for analysis of some otherwise really complicated problems.

Likewise, our analysis following local-to-zero asymptotic approximations focuses attention on the most relevant features of the problem and of the distribution of the measurement errors, and allows us to provide a practical approach for the settings where none was previously available.

The term "small" measurement error approximation is a misnomer. As shall be clear from the subsequent analysis, the corresponding bias can be larger than the incidental parameter bias (can asymptotically dominate). As we illustrate in the Monte Carlo experiments, the approach works well with rather large measurement errors that cause naive estimators to have large bias (sometimes much larger than the incidental parameter bias).

In Section 3, we begin with the analysis with a simpler, although sometimes also most practical case, given the data limitations often encountered. We then extend the estimation approach to handle serially correlated measurement errors, and measurement errors of larger magnitude.

Section **??** focuses on the choice of instruments. We show how the insights of GH86 can be applied in the nonlinear panel data settings.

Section 4 develops the large sample theory for the proposed estimators. The estimators are shown to be consistent, asymptotically normally distributed and asymptotically unbiased. The formulas for the estimation of the asymptotic variances are also provided.

Section 5 considers the settings in which the data comes from a large network. Such network data can be related to the large-T panel data. We show how the errors-in-variables bias can be addressed in the network settings, and how the network structure can provide instruments to identify the model.

The proofs are collected in the Appendix.

Notation. $E_T[A_t] \equiv \frac{1}{T}\sum_{t=1}^{T} A_t$, $E_{NT}[A_{it}] \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} A_{it}$. Also, let $E_i[A_{it}] \equiv \text{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} A_{it}$, i.e., $E_i[A_{it}]$ is the expectation w.r.t. the distribution of the i'th cross-section. We use $[k]$ to denote $1,\ldots,k$.

2 Biases of m-estimators in large-T panel data

Consider m-estimator of the form

$$\left(\widehat{\theta}^{\mathrm{m}}, \widehat{\alpha}_1^{\mathrm{m}}, \ldots, \widehat{\alpha}_N^{\mathrm{m}}\right) = \underset{\theta,\alpha_1,\ldots,\alpha_N}{\operatorname{argmax}} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\ell\left(X_{it}, S_{it}, \theta, \alpha_i\right). \tag{1}$$

where $\theta \in \mathbb{R}^{d_\theta}$ and $\alpha_i \in \mathbb{R}^{d_\alpha}$. Here X_{it} are the mismeasured covariates, while S_{it} collects all other observed variables. The primary example is the (quasi-)MLE estimator with $\ell \equiv \log f$. We use superscript "m" to denote the m-estimators.

In this section we obtain bias formula for such estimators. In the following sections we introduce estimators that are robust to the presence of measurement errors.

We assume that the true parameters solve the population problem

$$(\theta_0, \alpha_{i0}) = \underset{\theta,\alpha_i}{\operatorname{argmax}}\, E\left[\ell\left(X_{it}^*, S_{it}, \theta, \alpha_i\right)\right]. \tag{2}$$

Example 1 *Suppose Y_{it} is binary, and*

$$E\left[Y_{it}|X_{it}^*, W_{it}\right] = \Lambda\left(\alpha_{i0} + \theta_{0,1}' X_{it}^* + \theta_{0,W}' W_{it}\right),$$

where Λ is the CDF of a continuous distribution, such as Logistic or Normal. Then $\theta_0 = (\theta_{0,1}, \theta_{0,W})$, $S_{it} = (Y_{it}, W_{it})$, and

$$\ell\left(X_{it}, S_{it}, \theta, \alpha_i\right) = Y_{it}\log\Lambda\left(\alpha_i + \theta_1' X_{it}^* + \theta_W' W_{it}\right) + \left(1 - Y_{it}\right)\log\left(1 - \Lambda\left(\alpha_i + \theta_1' X_{it}^* + \theta_W' W_{it}\right)\right).$$

Panel Probit model corresponds to taking Λ to be the CDF Φ of Normal distribution. Dynamic models are included, for example, one can include lagged outcomes $Y_{i,t-1}$ as a part of vector W_{it}.

Example 2 *Linear model with individual-specific slopes and intercepts*

$$Y_{it} = \gamma_{i0} + \beta_{i0}' X_{it}^* + \tau_0' W_{it} + U_{it}, \qquad E\left[U_{it}|X_{it}^*, W_{it}\right] = 0.$$

Here $\theta_0 = (\tau_0, \sigma_{U_{it}}^2)$ *and* $\alpha_{i0} \equiv (\gamma_{i0}, \beta_{i0})$. *Then, Gaussian (quasi) MLE estimator corresponds to*

$$\ell\left(X_{it}, S_{it}, \tau, \sigma_U, \gamma_i, \beta_i\right) = -\frac{1}{2\sigma_U^2}\left(\gamma_i + \beta_i' X_{it} + \tau' W_{it} - Y_{it}\right)^2 - \frac{1}{2}\ln\left(\sigma_U^2\right).$$

Example 3 *Panel nonlinear least-squares model*

$$E\left[Y_{it}|X_{it}^*, \alpha_i\right] = m\left(X_{it}^*, W_{it}, \theta, \alpha_i\right)$$

can be estimated using

$$\ell\left(X_{it}, S_{it}, \tau, \sigma_U, \gamma_i, \beta_i\right) = -\left(m\left(X_{it}, W_{it}, \theta, \alpha_i\right) - Y_{it}\right)^2.$$

To simplify the exposition, in this section we assume that X_{it} and α_i are scalar, although the results naturally extend to multivariate X_{it} and α_i.

We assume that

$$\begin{aligned} X_{it} &= X_{it}^* + \varepsilon_{it}, \text{ where} \qquad (3)\\ E\left[\varepsilon_{it}|X_{it}^*, S_{it}\right] &= 0,\ E\left[\varepsilon_{it}^2|X_{it}^*, S_{it}\right] = \sigma^2. \qquad (4)\end{aligned}$$

The above equations have the spirit of the "classical" measurement error assumption, but are weaker, since they do not rule out dependence between ε_{it} and (X_{it}^*, S_{it}).

Under some regularity conditions,

$$\sqrt{NT}\left(\widehat{\theta}^{\mathrm{m}} - \theta_0 - \mathcal{B}_{\sigma,N,T}\right) \to_d N\left(0, \Omega\right), \qquad (5)$$

where $\mathcal{B}_{\sigma,N,T}$ is the asymptotic bias of $\widehat{\theta}^{\mathrm{m}}$. To present the expressions for the asymptotic bias of $\widehat{\theta}^{\mathrm{m}}$ we need to introduce some additional notation. Let

$$u_{it}\left(\theta, \alpha\right) \equiv \nabla_\theta \ell\left(X_{it}, S_{it}, \theta, \alpha\right), \qquad v_{it}\left(\theta, \alpha\right) \equiv \nabla_{\alpha_i} \ell\left(X_{it}, S_{it}, \theta, \alpha\right).$$

In this section we use superscripts denote partial derivatives with respect to parameters θ and α, e.g., $u_{it}^\alpha\left(\theta, \alpha_i\right) \equiv \nabla_\alpha u\left(X_{it}, S_{it}, \theta, \alpha_i\right)$. We omit the arguments when a function is evaluated at (θ_0, α_{i0}), e.g., of $u_{it}^\alpha \equiv u_{it}^\alpha\left(\theta_0, \alpha_{i0}\right)$. We put $*$ in the superscript to indicate evaluating a function at X_{it}^* rather than X_{it}, e.g., $u_{it}^{*\alpha} \equiv \nabla_\alpha u\left(X_{it}^*, S_{it}, \theta_0, \alpha_{i0}\right)$. We use subscripts x to denote derivatives with respect to x, e.g., $u_{itx} \equiv \nabla_x u_{it}$, $u_{itxx} \equiv \nabla_{xx} u_{it}$, and $u_{itxxx} \equiv \nabla_{xxx} u_{it}$. We also write $u_{itx}^{(k)}$ to denote $\left(\partial^k/\partial x^k\right) u_{it}$. Finally, we let $E_i[\cdot]$ denote expectation with respect to the distribution of the i'th cross-section.

To present the expressions for the asymptotic bias of $\widehat{\theta}^{\mathrm{m}}$ we need to introduce some additional notation. Let

$$u_{it}(\theta,\alpha) \equiv \frac{\partial}{\partial\theta}\ell(X_{it},S_{it},\theta,\alpha), \qquad v_{it}(\theta,\alpha) \equiv \frac{\partial}{\partial\alpha_i}\ell(X_{it},S_{it},\theta,\alpha).$$

Let additional subscripts denote partial derivatives, e.g., $u_{it\theta}(\theta,\alpha) = \partial u_{it}(\theta,\alpha)/\partial\theta$. Let $E_i[\cdot]$ denote expectation w.r.t. the distribution of the i'th cross-section. Let

$$\begin{aligned}
&v_{it} \equiv v_{it}^*\left(\overline{\theta},\overline{\alpha}_i\right), \qquad v_{it}^{*\alpha} \equiv \nabla_{\alpha_i} v_{it}\left(\overline{\theta},\overline{\alpha}_i\right), \qquad v_{it}^{*\alpha\alpha} \equiv \nabla_{\alpha_i\alpha_i} v_{it}\left(\overline{\theta},\overline{\alpha}_i\right),\\
&u_{it} \equiv u_{it}\left(\overline{\theta},\overline{\alpha}_i\right), \qquad u_{it}^{*\alpha} \equiv \nabla_{\alpha_i} u_{it}\left(\overline{\theta},\overline{\alpha}_i\right), \qquad u_{it}^{*\alpha\alpha} \equiv \nabla_{\alpha_i\alpha_i} u_{it}\left(\overline{\theta},\overline{\alpha}_i\right),\\
&\psi_{it} \equiv -E_i\left[v_{it}^{*\alpha}\right]^{-1} v_{it}^*.
\end{aligned}$$

Assumption 1 *$\sigma^2 = o\left((NT)^{-\omega}\right)$ for some $\omega > 1/3$ and $E\left[\left|\varepsilon_{it}/\sigma\right|^k\right]$ is bounded for all $k \geq 1$.*

Assumption 2 *For some constants $\delta > 0$ and C, $\left|\frac{\partial^{k+l+m}}{\partial\theta^k\partial\alpha^l\partial x^m}\ell(x,s,\theta,\alpha)\right| \leq C$ for all $\theta \in B_{(\theta_0)}(\delta)$ and all α, x, s, and nonnegative integers k, l, m such that $k+l \leq 5$, $m \leq 3$.*

Condition that all moments of $|\varepsilon_t/\sigma|$ are bounded can be relaxed. Note that we are imposing assumptions on the true distribution of the data (X_{it}^*, S_{it}). The nonstandard asymptotic approximation with $\sigma^2 \to 0$ we employ only applies to the modelling of the measurement error. Some of the results make use of the following standard Assumption.

Assumption 3 *$N/T \to \kappa^2 \in (0,\infty)$ as $N \to \infty$, $T \to \infty$.*

Proposition 4 *Suppose Assumptions 3,1,2, and* **??** *hold. Then in equation (5)*

$$\mathcal{B}_{\sigma,N,T} = \sigma^2 B_{\mathrm{ME}} + \frac{1}{T} B_{\mathrm{INC}} + O\left(\sigma^3 + \frac{1}{T^2} + \frac{\sigma^2}{T}\right),$$

where

$$\begin{aligned}
B_{\mathrm{ME}} &\equiv H^* E\left[u_{it}^{*\alpha}\beta_{ME,i} + \frac{1}{2}u_{itxx}^*\right], &(6)\\
B_{\mathrm{INC}} &\equiv H^* E\left[u_{it}^{*\alpha}\beta_{INC,i} + \mathbb{C}_{LR,i}\left[u_{it}^{*\alpha},\psi_{it}^*\right] + \frac{1}{2}u_{it}^{*\alpha\alpha}V_{LR,i}\left[\psi_{it}^*\right]\right], &(7)\\
\psi_{it}^* &\equiv Q_i^* v_{it}^*, \qquad Q_i^* \equiv -E_i\left[v_{it}^{*\alpha}\right]^{-1},\\
\beta_{ME,i} &\equiv \frac{1}{2}Q_i^* E_i\left[v_{itxx}^*\right],\\
\beta_{INC,i} &\equiv Q_i^*\left(\mathbb{C}_{LR,i}\left[v_{it}^{*\alpha},\psi_{it}^*\right] + \frac{1}{2}E_i\left[v_{it}^{*\alpha\alpha}\right]V_{LR,i}\left[\psi_{it}^*\right]\right).
\end{aligned}$$

$$H^{*} \equiv -\left(E\left[u_{it}^{*\theta} - u_{it}^{*\alpha} \cdot E_i\left[v_{it}^{*\theta}\right]/E_i\left[v_{it}^{*\alpha}\right]\right]\right)^{-1}. \tag{8}$$

The bias of $\widehat{\theta}^{\mathrm{m}}$ consists of three parts: the measurement error bias $\sigma^2 B_{\mathrm{ME}}$, the incidental parameter bias $\frac{1}{T}B_{\mathrm{INC}}$, and the higher order bias.

The measurement error bias can be decomposed into two parts. The first term in equation (6) comes from the errors-in-variables bias $\beta_{\mathrm{ME},i}$ in estimation of the fixed-effects. The second term arises directly from the effect of the measurement error on the u_{it}, which is nonlinear in X_{it}.

The incidental parameter bias formula is the same as in HN04 and HK11. As HN04 explain, the incidental parameter bias has three sources. The first term in equation (7) is due to the bias $\beta_{\mathrm{INC},i}$ of the nuisance parameters α_i. The second term is due to the parameters θ and α_i being estimated from the same data. The third term is the usual nonlinearity bias, due to the randomness in $\widehat{\alpha}_i$.

It is worth briefly discussing the higher order bias, which is $O\left(\sigma^3 + 1/T^2 + \sigma^2/T\right)$. The first two terms are the next order terms of measurement error and incidental parameter biases, respectively. Term $O\left(\sigma^2/T\right)$ represents the effect of measurement error on the first order incidental parameter bias.

To illustrate these points, consider a simplified version of Example 2:

$$Y_{it} = \alpha_{i0}X_{it}^{*} + U_{it}, \qquad E\left[U_{it}|X_i^{*}\right] = 0, \qquad \theta_0 \equiv E\left[U_{it}^2\right]. \tag{9}$$

Then the Gaussian (pseudo)-likelihood function is $E_{NT}\left[\ell_{it}\left(\theta,\alpha_i\right)\right]$, where

$$\begin{aligned}
\ell_{it}\left(\theta,\alpha_i\right) &\equiv \log f_{it}\left(\theta,\alpha_i\right) = -\frac{1}{2\theta}\left(\alpha_i X_{it} - Y_{it}\right)^2 - \frac{1}{2}\ln\theta, \\
u_{it}\left(\theta,\alpha_i\right) &= \frac{1}{2\theta^2}\left(\left(\alpha_i X_{it} - Y_{it}\right)^2 - \theta\right), \qquad v_{it}\left(\theta,\alpha_i\right) = -\frac{1}{\theta}\left(\alpha_i X_{it} - Y_{it}\right)X_{it},
\end{aligned}$$

and the MLE estimators are

$$\hat{\alpha}_i \equiv E_T\left[X_{it}Y_{it}\right]/E_T\left[X_{it}^2\right] \ \text{ and } \ \hat{\theta} \equiv E_N E_T\left[\left(\hat{\alpha}_i X_{it} - Y_{it}\right)^2\right].$$

$$
\begin{aligned}
v_{it}^{*\alpha} &= -\frac{X_{it}^{*2}}{\theta_0}, \qquad v_{it}^{*\alpha\alpha} = 0, \qquad Q_i^* = \frac{\theta_0}{E_i\left[X_{it}^{*2}\right]}, \qquad \psi_{it}^* = \frac{1}{E_i\left[X_{it}^{*2}\right]}\left(Y_{it} - \alpha_{i0}X_{it}^*\right)X_{it}^*, \\
u_{it}^{*\alpha} &= \frac{1}{\theta_0^2}X_{it}^*\left(\alpha_{i0}X_{it}^* - Y_{it}\right), \qquad E_i\left[u_{it}^{*\alpha}\right] = 0, \qquad u_{it}^{*\alpha\alpha} = \frac{1}{\theta_0^2}X_{it}^{*2} \\
u_{it}^{*\theta} &= -\frac{1}{\theta_0^3}\left(\alpha_{i0}X_{it}^* - Y_{it}\right)^2 + \frac{1}{2\theta_0^2}, \qquad H^* = 2\theta_0^2, \qquad u_{itxx}^* = \frac{\alpha_{i0}^2}{\theta_0^2}, \text{ hence} \\
\beta_{\mathrm{ME},i} &= -\frac{\alpha_i}{E_i\left[X_{it}^{*2}\right]} \\
B_{\mathrm{INC}} &= -\theta_0, \qquad B_{\mathrm{ME}} = E\left[\alpha_{i0}^2\right], \text{ i.e.,}
\end{aligned}
$$

$$
\sqrt{NT}\left(\hat{\theta} - \left\{\theta_0 - \frac{1}{T}\theta_0 + \sigma^2 E\left[\alpha_{i0}^2\right]\right\}\right) \to_d N\left(0, \Omega_\ell\right).
$$

Remark 1 *In this example the incidental parameter bias and the measurement error bias have the opposite signs. In such cases, an estimator that only corrects the incidental parameter bias, but ignores the errors-in-variables bias, usually has larger bias than the naive estimator that ignores the incidental parameter problem.*

Remark 2 *The incidental parameter bias $B_{\mathrm{INC}}/T = -\theta_0/T$ is identical to the incidental parameter bias in the Neyman and Scott (1948) model*

$$
Y_{it} = \alpha_i + U_{it}, \qquad E\left[U_{it}\right] = 0, \qquad \theta_0 \equiv E\left[U_{it}^2\right]. \tag{10}
$$

Remark 3 *Since the model of equation (9) is very simple, one can directly calculate higher order terms of the asymptotic bias of $\hat{\theta}$ to be*

$$
B_{T,\sigma} = -\theta_0 \cdot \frac{1}{T} + \sigma^2 \cdot \sum_{k=0}^{K/2} E\left[\alpha_i^2\left(-\frac{\sigma^2}{\sigma_{X_{it}^*}^2}\right)^k\right] + \frac{\sigma^2}{T} \cdot E\left[-\alpha_i^2\left(\frac{1}{\sigma_{X_{it}^*}^2} + 1\right)\right] + o\left(\frac{\sigma^2}{T}\right) + O\left(\sigma^{K+1} + \frac{1}{T^2}\right),
$$

where we have assumed that $E\left[X_{it}^*\right] = 0$ and K is even.

Remark 4 *We can continue the MME expansion to a higher order and obtain the expression of the form $\overline{\theta} = \theta_0 + \sum_{k=2}^{K}\sigma^k B_{\sigma,k}^* + O\left(\sigma^{K+1}\right)$. The leading terms in the expansion of $\overline{B}$ remain $B^* + \frac{\sigma^2}{T}C_B$, so equation (**??**) becomes*

$$
\sqrt{NT}\left(\widehat{\theta}^m - \theta_0 - \left[\frac{1}{T}B^* + \sum_{k=2}^{K}\sigma^k B_{\sigma,k}^* + \frac{\sigma^2}{T}C_B + O\left(\sigma^{K+1} + \frac{\sigma^3}{T} + \frac{1}{T^2}\right)\right]\right) \to_d N\left(0, \Omega_{HN04}^*\right).
$$

Under what conditions can we ignore the bias correction $\frac{\sigma^2}{T}C_B$ of the incidental parameter bias? The remainder of order σ^{K+1} can be ignored if we make an assumption *that as $\sigma^2 = o\left((NT)^{-1/(K+1)}\right)$. Hence, $\sqrt{NT}\frac{\sigma^2}{T}$ is negligible as long as $\frac{1}{T}(NT)^{\frac{1}{2}-\frac{1}{K+1}}$ is bounded, i.e., $N = O\left(T^{\frac{K+3}{K-1}}\right)$. Thus, bias correcting the incidental parameter bias term is not necessary as long as $K \leq 3$. When $K \geq 4$ would not be needed only if the condition $N = o(T^3)$ strengthened to $N = O\left(T^{\frac{K+3}{K-1}}\right)$. Intuitively, when T is small relative to N, incidental parameter bias correction plays a more important role, and needs to be corrected in the presence of larger measurement error. For example, if $K = 4$ one needs to either assume that $N/T^{7/3}$ is bounded, or estimate the term $\frac{\sigma^2}{T}C_B$ and include it as a part of the bias correction.*

Remark 5 *Few papers study large-T panel data settings allowing $T = o(N)$ in the asymptotic approximations; the majority of the literature develops asymptotic theory only under the assumption $N/T \to \kappa^2 \in (0, \infty)$, since allowing $T = o(N)$ substantially complicates the theoretical analysis. Under this asymptotic approximation, term $\frac{\sigma^2}{T}C_B$ in the above expressions can always be ignored. However, in practice T is often much smaller than N, and hence the conclusion of the previous Remark has practical implications regardless of the choice of the asymptotic framework.*

To correct the incidental parameter bias, one can use a panel jackknife procedure. For instance, following HN04 let $\widehat{\theta}$ denote an estimator that uses full dataset, and define $\widehat{\theta}_{(t)}$ to be the estimator obtained applying the estimator to the dataset that excludes observations from t'th time-period. NH04 show that estimator

$$\widehat{\theta}_J \equiv T\widehat{\theta} - (T-1)\frac{1}{T}\sum_{t=1}^{T}\widehat{\theta}_{(t)} \tag{11}$$

removes the incidental parameter bias of order $\frac{1}{T}$ in static models. For dynamic panels the jackknife procedure of Dhaene and Jochmans (2015) can be used.

It is important to note that jackknifing does not affect the measurement error bias $\sigma^2 B_\sigma^*$. To see this, let us denote the bias of the estimator $\widehat{\theta}^{\mathrm{m}}$ that uses all T time periods by

$$\mathcal{B}_{T,\sigma} \equiv \frac{1}{T}B^* + \sum_{k=2}^{K}\sigma^k B_{\sigma,k}^* + \frac{\sigma^2}{T}C_B + O\left(\sigma^{K+1} + \frac{\sigma^3}{T} + \frac{1}{T^2}\right).$$

Then the bias of $\widehat{\theta}_J^{\mathrm{m}}$ is approximately

$$T\mathcal{B}_{T,\sigma} - (T-1)\mathcal{B}_{T-1,\sigma} = \sum_{k=2}^{K}\sigma^k B_{\sigma,k}^* + O\left(\sigma^{K+1} + \frac{1+\sigma^2}{T^2}\right).$$

Thus, jackknife is able to remove two sources of bias $\frac{1}{T}B^*$ but leaves the leading measurement error bias term untouched. An advantage of the jackknife bias correction is that it and $\frac{\sigma^2}{T}C_B$,

Let us illustrate the two sources of EIV bias of the fixed effect estimators of common parameter θ_0 in two very simple examples.

Example 4 *"Textbook" model:*

$$Y_{it} = \alpha_i + \theta X_{it} + U_{it}.$$

Here $u_{xx} = -2\theta$, $v_{xx} = 0$, so there is no "direct" errors-in-variables bias in $\widehat{\alpha}_i$, the errors-in-variables bias of $\widehat{\theta}$ comes from the bias in θ, with $B_{\text{ME}} = -\theta_0/\sigma^2_{X^*}$, which is an expression familiar from standard OLS regression without any incidental parameters.

Example 5 *Mirror of the "textbook" model:*

$$Y_{it} = \theta + \alpha_i X_{it} + U_{it}$$

Here $u_{xx} = 0$, $v_{xx} = -2\alpha_i$, so the errors-in-variables bias of $\widehat{\theta}$ is solely attributed to the bias of $\widehat{\alpha}_i$, with $B_{\text{ME}} = -E\left[\alpha_i X^*_{it}\right]/\sigma^2_{X^*}$.

The key ingredient of the bias expressions above is the variance of the measurement error σ^2. If σ^2 is known, or can be estimated from a validation dataset or repeated measurements, the above expressions can be used to bias correct $\widehat{\theta}^{\text{m}}$ to obtain asymptotically unbiased estimators of θ_0. However, such additional data allowing direct estimation of σ^2 is rarely available in practice. In the next section we develop an estimation approach in which parameters θ, α_i, and σ^2 are jointly estimated by using lagged covariates as instruments to handle the problem of measurement errors.

3 Estimation

3.1 Motivation

As pointed out by GH86, in panel data models, lagged values of mismeasured covariates provide instrumental variables that can be used to address the errors-in-variables problem. To make use of such instruments, we need to set the problem in the more general moment conditions framework.

Consider the individual-specific moment conditions

$$E_i\left[g\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]=0 \qquad \text{if and only if} \qquad (\theta, \alpha_i)=(\theta_0, \alpha_{i0}) \;\; \forall i \in [N], \tag{12}$$

where $d_g \equiv \dim(g) \geq d_\theta + d_\alpha$, and the moment condition may potentially include a vector of instruments Z_{it}. For instance, in Example 1 one may take

$$g\left(X_{it}, Y_{it}, W_{it}, Z_{it}, \theta, \alpha_i\right)=\left(Y_{it}-\Lambda\left(\theta_1 X_{it}+\theta_W' W_{it}+\alpha_i\right)\right) h\left(X_{it}, W_{it}, Z_{it}, \theta, \alpha_i\right),$$

where $h\left(X_{it}, W_{it}, Z_{it}, \theta, \alpha_i\right)$ is a vector of functions of $\left(X_{it}, W_{it}, Z_{it}\right)$ that in general could depend on (θ, α_i). In particular, taking

$$h\left(X_{it}, W_{it}, Z_{it}, \theta, \alpha_i\right)=\frac{\Lambda'\left(\theta_1 X_{it}+\theta_W' W_{it}+\alpha_i\right)}{\left(1-\Lambda\left(\theta_1 X_{it}+\theta_W' W_{it}+\alpha_i\right)\right) \Lambda\left(\theta_1 X_{it}+\theta_W' W_{it}+\alpha_i\right)}\begin{pmatrix} X_{it} \\ W_{it} \\ 1 \end{pmatrix},$$

corresponds to the MLE estimator.

When covariates are mismeasured, generally $E_i\left[g\left(X_{it}, S_{it}, Z_{it}, \theta_0, \alpha_{i0}\right)\right] \neq 0$. To address this problem it is usually easy to introduce additional moment conditions that make use of the instrumental variables. For instance, in Example 1 the moment condition

$$E_i\left[\left(Y_{it}-\Lambda\left(\theta_{0,1} X_{it}^*+\theta_{0,W}' W_{it}+\alpha_{i0}\right)\right) Z_{it}\right]=0, \tag{13}$$

where $Z_{it}=X_{i,\tau}$ for any $\tau<t$, or, more generally, $Z_{it}=\varphi\left(X_{i,\tau}\right)$ for some vector of functions $\varphi(\cdot)$.

It is important to note that the expectation in equation (13) involves X_{it}^* and not X_{it}. Unless Λ is a linear function, the instrument itself does not resolve the problem of errors-in-variables, and generally $E_i\left[\left(Y_{it}-\Lambda\left(\theta_{0,1} X_{it}+\theta_{0,W}' W_{it}+\alpha_{i0}\right)\right) Z_{it}\right] \neq 0$. As pointed out by Amemiya (1985), in nonlinear models, instrumental variables do not alleviate errors-in-variables bias by themselves.

To handle the problem of measurement errors we make use of Assumption 1 and consider the following expansion of the moment condition:

$$\begin{aligned}
& E_i\left[g\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] \\
= & E_i\left[g\left(X_{it}^*+\varepsilon_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] \\
= & E_i\left[g\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]+E_i\left[\varepsilon_{it} g_x\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] \\
& +E_i\left[\frac{1}{2} \varepsilon_{it}^2 g_{xx}\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]+O\left(E_i\left[\left|\varepsilon_{it}\right|^3\right]\right) \\
= & E_i\left[g\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]+0+\frac{1}{2} \sigma^2 E_i\left[g_{xx}\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]+O\left(\sigma^3\right).
\end{aligned} \tag{14}$$

By Assumption 1, $O(\sigma^3) = o\left((NT)^{-1/2}\right)$ and is negligible. Evaluating the above at (θ_0, α_{i0}) and using equation (12) we obtain

$$E_i\left[g\left(X_{it}, S_{it}, Z_{it}, \theta_0, \alpha_{i0}\right)\right] = \frac{1}{2}\sigma^2 E_i\left[g_{xx}\left(X_{it}^*, S_{it}, Z_{it}, \theta_0, \alpha_{i0}\right)\right] + o\left((NT)^{-1/2}\right).$$

Since the expectation of the moment condition is not zero at the true parameters, the estimator based on the moment conditions g_{it} is biased. One can show that the measurement error induces bias of order σ^2. In particular, if $\sqrt{NT}\sigma^2 \to \infty$, like MLE, the GMM estimators will not be $\sqrt{NT}$ consistent.

Had we known the value of the second term on the last line of equation (14), we could have corrected the moment condition:

$$E_i\left[g\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] - \frac{1}{2}\sigma^2 E_i\left[g_{xx}\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] = E_i\left[g\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] + o\left((NT)^{-1/2}\right).$$

Above, on the right hand side we have the moment condition evaluated at X_{it}^*, i.e., free from the measurement error bias. Of course, the above correction is infeasible, because we do not know σ^2 or $E_i\left[g_{xx}\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]$, which depends on the unobserved X_{it}^*. As Evdokimov and Zeleneev (2016) point out, as long as function g is sufficiently smooth, we can approximate the correction term by

$$\frac{1}{2}\sigma^2 E_i\left[g_{xx}\left(X_{it}^*, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] = \frac{1}{2}\sigma^2 E_i\left[g_{xx}\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right)\right] + O\left(\sigma^4\right).$$

We can estimate $E_i\left[g_{xx}\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right)\right]$ by $\frac{1}{T}\sum_{t=1}^{T} g_{xx}\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right)$, so the only unknown part of the correction is σ^2.

We introduce an additional parameter $\gamma \in \mathbb{R}$ and the following corrected moment conditions:

$$\psi_{it}\left(\theta, \alpha_i, \gamma\right) \equiv g_{it}\left(\theta, \alpha_i\right) - \gamma g_{it,xx}\left(\theta, \alpha_i\right) \equiv g\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right) - \gamma g_{xx}\left(X_{it}, S_{it}, Z_{it}, \theta, \alpha_i\right). \tag{15}$$

Let $\gamma_0 \equiv \sigma^2/2$. Then

$$E_i\left[\psi_{it}\left(\theta_0, \alpha_{i0}, \gamma_0\right)\right] = O\left(\sigma^3\right) = o\left((NT)^{-1/2}\right). \tag{16}$$

The moment conditions ψ_{it} have zero mean (up to the negligible term) at the true parameter values, and hence can be used to obtain asymptotically unbiased estimators, since we can estimate $E_i\left[\psi_{it}\left(\theta, \alpha_i, \gamma\right)\right]$ by $E_T\left[\psi_{it}\left(\theta, \alpha_i, \gamma\right)\right]$.

Remark 6 *If one assumes that ε_{it} has zero skewness (e.g., ε_{it} is symmetric), equation (16) holds with $O\left(\sigma^4\right)$ instead of $O\left(\sigma^3\right)$ on the right-hand side.*

Remark 7 *The derivative g_{xx} can be computed numerically, hence the corrected moment conditions can be computed by an automatic "black-box" procedure without requiring any additional programming from the researcher.*

3.2 Implementation

Let $\overline{\psi}_i(\theta, \alpha_i, \gamma) \equiv \frac{1}{T}\sum_{t=1}^{T} \psi_{it}(\theta, \alpha_i, \gamma)$ and let $\widehat{\Xi}_i$ be some some p.s.d. weighting matrices. Consider the following estimator

$$\left(\widehat{\theta}, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_N, \widehat{\gamma}\right) = \underset{(\theta,\alpha_1,\ldots,\alpha_N,\gamma)\in\Theta\times\mathcal{A}^N\times\Gamma}{\operatorname{argmin}} \sum_{i=1}^{N} \widehat{Q}_i(\theta, \alpha_i, \gamma), \text{ where} \tag{17}$$
$$\widehat{Q}_i(\theta, \alpha_i, \gamma) \equiv \overline{\psi}_i(\theta, \alpha_i, \gamma)' \, \widehat{\Xi}_i \overline{\psi}_i(\theta, \alpha_i, \gamma).$$

The properties of such GMM estimators have been analyzed by FL13, who show that the estimators of common parameters are $\sqrt{NT}$ consistent, but suffer from the incidental parameter bias of order $\frac{1}{T}$. This bias is linked to the analysis of higher-order properties of GMM estimators by Newey and Smith (2004). This bias can be corrected by either a jackknife procedure such as in equation (11), or by the analytical bias correction developed by FL13.

We can take the weighting matrices $\widehat{\Xi}_i$ to be sample estimators of the (long-run) variance-covariance matrices of $\psi_{it}\left(\widetilde{\theta}, \widetilde{\alpha}_i\right)$ evaluated at some preliminary parameter values $\left(\widetilde{\theta}, \widetilde{\alpha}_i\right)$. The (biased) m-estimator (1) can be used as preliminary parameter values for estimation of $\widehat{\Xi}_i$. One can also use an incidental parameter bias corrected version of this estimator, e.g., jackknife bias corrected estimator.

One advantage that m-estimators (1) often have is possessing a globally convex criterion function, which greatly simplifies solving the high dimensional optimization problem. Bias corrected criterion functions usually do not have this property, so, in practice, it is important to use good starting values in the optimization problem (17). For example, the (biased) m-estimator (1) provide very good starting values for the optimization problem (17). From the standpoint of numerical optimization, the speed and reliability of numerical solvers of (17) can be greatly improved if one provides a routine computing analytical gradient of the criterion function, and makes use of the sparsity of the Hessian in this problem (note that even though the Hessian matrix has $O\left(N^2\right)$ elements, the number of its nonzero elements grows linearly with N.) In Appendix **??** we provide additional details on the computation of these estimators.

Then, for any (θ,γ) and i, $\widehat{\alpha}_i(\theta,\gamma)$ solve

$$FOC_{\alpha_i}: 0=\overline{\Psi}_{i\alpha_i}\left(\theta,\widehat{\alpha}_i(\theta,\gamma),\gamma\right)'\widehat{\Xi}_i\overline{\psi}_i\left(\theta,\widehat{\alpha}_i(\theta,\gamma),\gamma\right)$$

and $\left(\widehat{\theta},\widehat{\gamma}\right)$ solve

$$\begin{aligned}
FOC_{\theta} &: 0=\sum_{i=1}^{N}\overline{\Psi}_{i\theta}\left(\theta,\widehat{\alpha}_i(\theta,\gamma),\gamma\right)'\widehat{\Xi}_i\overline{\psi}_i\left(\theta,\widehat{\alpha}_i(\theta,\gamma),\gamma\right),\\
FOC_{\gamma} &: 0=\sum_{i=1}^{N}\overline{\Psi}_{i\gamma}\left(\theta,\widehat{\alpha}_i(\theta,\gamma),\gamma\right)'\widehat{\Xi}_i\overline{\psi}_i\left(\theta,\widehat{\alpha}_i(\theta,\gamma),\gamma\right)
\end{aligned}$$

Then

$$\begin{aligned}
0 &\approx \sum_{i=1}^{N}\overline{\Psi}_{i\theta}\left(\theta_0,\widehat{\alpha}_i(\theta_0),\gamma_0\right)'\widehat{\Xi}_i\left\{\overline{\psi}_i\left(\theta_0,\widehat{\alpha}_i(\theta_0),\gamma_0\right)+\overline{\Psi}_{i\theta}\left(\theta_0,\widehat{\alpha}_i(\theta_0),\gamma_0\right)\left(\widehat{\theta}-\theta_0\right)\right\},\\
\sqrt{NT}\left(\widehat{\theta}-\theta_0\right) &= \left\{E_N\left[\overline{\Psi}_{i\theta}\left(\theta_0,\widehat{\alpha}_i(\theta_0,\gamma_0),\gamma_0\right)'\widehat{\Xi}_i\overline{\Psi}_{i\theta}\left(\theta_0,\widehat{\alpha}_i(\theta_0,\gamma_0),\gamma_0\right)\right]\right\}\\
&\quad\times\sqrt{NT}E_N\left[\overline{\Psi}_{i\theta}\left(\theta_0,\widehat{\alpha}_i(\theta_0,\gamma_0),\gamma_0\right)'\widehat{\Xi}_i\overline{\psi}_i\left(\theta_0,\widehat{\alpha}_i(\theta_0,\gamma_0),\gamma_0\right)\right].
\end{aligned}$$

Here

$$\begin{aligned}
\overline{\psi}_i\left(\theta_0,\widehat{\alpha}_i(\theta_0,\gamma_0),\gamma_0\right) &\approx \overline{\psi}_i\left(\theta_0,\alpha_{i0},\gamma_0\right)+\overline{\Psi}_{i\alpha_i}\left(\theta_0,\alpha_{i0},\gamma_0\right)\left(\widehat{\alpha}_i(\theta_0,\gamma_0)-\alpha_{i0}\right)\\
&\quad+\overline{\Psi}_{i\alpha_i\alpha_i}\left(\theta_0,\alpha_{i0},\gamma_0\right)\left(\widehat{\alpha}_i(\theta_0,\gamma_0)-\alpha_{i0}\right)^2.
\end{aligned}$$

Under some regularity conditions (see Section 4))

$$\sqrt{T}\overline{\psi}_i\left(\theta_0,\widehat{\alpha}_i(\theta_0,\gamma_0),\gamma_0\right)=\sqrt{T}\overline{\psi}_i\left(\theta_0,\alpha_{i0},\gamma_0\right)+T^{-1/2}Q_{1i}+T^{-1}R_{2i},$$

where $\sup_{i\in[N]}|R_{2i}|=o_P\left(\sqrt{T}\right)$.

3.3 Larger Measurement Errors

When the measurement error variance are moderately large (relative to the sample size and the degree of nonlinearity of the model), one may want to continue the Taylor expansion of the moment condition (14) to a higher order, and then correct the moment conditions similarly to equation (15). As Evdokimov and Zeleneev (2016) point out, a naive application of this strategy does not work, and one needs to tweak the definition of the corrected moment conditions ψ_{it}.

$$g_{it}=g_{it}^{*}+\varepsilon_{it}g_{it}^{*x}+\frac{1}{2}\varepsilon_{it}^{2}g_{it}^{*xx}+\frac{1}{6}\varepsilon_{it}^{3}g_{it}^{*xxx}+\frac{1}{24}\varepsilon_{it}^{4}g_{it}^{*xxxx}+\varepsilon_{it}^{5}\widetilde{g}_{it}^{xxxxx}$$

Let $\kappa_j \equiv \frac{1}{j!}E\left[\varepsilon_{it}^j\right]$ for $j = 2, 3, 4$. By analogy with equation (15), one may have hoped that the expectation of

$$\tilde{\psi}_{it} \equiv g_{it} - \kappa_2 g_{it}^{xx} - \kappa_3 g_{it}^{xxx} - \kappa_4 g_{it}^{xxxx} \tag{18}$$

is of order $o(\sigma^4)$. This turns out not to be the case, in fact

$$E\left[\tilde{\psi}_{it}\right] = O\left(\sigma^2\right).$$

It turns out that the measurement error correction in the definition of $\tilde{\psi}_{it}$ is insufficient. The crux is that we need to correct the correction terms in equation (18). The correction term g_{it}^{xx} has the property that $E\left[g_{it}^{xx}\right] = E\left[g_{it}^{*xx}\right] + O\left(\sigma^3\right)$, and hence itself needs to be corrected. To estimate $\kappa_2 E\left[g_{it}^{*xx}\right]$ one can take $\kappa_2 E\left[g_{it}^{xx}\right] - \kappa_2^2 E\left[g_{it}^{xxxx}\right] = \kappa_2 E\left[g_{it}^{*xx}\right] + O\left(\sigma^5\right)$. Thus, the valid corrected moment condition is defined as

$$\psi_{it}\left(\theta, \alpha_i, \kappa_2, \kappa_3, \kappa_4\right) \equiv g_{it}\left(\theta, \alpha_i\right) - \kappa_2 g_{it}^{xx}\left(\theta, \alpha_i\right) - \kappa_3 g_{it}^{xxx}\left(\theta, \alpha_i\right) - \left(\kappa_4 - \kappa_2^2\right) g_{it}^{xxxx}\left(\theta, \alpha_i\right).$$

Remark 8 *We can consider a "reduced form" version of the moment condition $\psi_{it}\left(\theta, \alpha_i, \kappa_2, \kappa_3, \kappa_4\right)$, and estimate parameters $\theta, \alpha_i, \gamma_1, \gamma_2, \gamma_3$ using the moment condition*

$$\psi_{it}\left(\theta, \alpha_i, \gamma_1, \gamma_2, \gamma_3\right) \equiv g_{it}\left(\theta, \alpha_i\right) - \gamma_1 g_{it}^{xx} - \gamma_2 g_{it}^{xxx} - \gamma_3 g_{it}^{xxxx}.$$

This provides valid estimators of and θ_0, α_{i0}, and the moments $\{\kappa_j\}$ can be deduced from the estimated γ.

If one is interested in imposing basic restrictions on the moment relationships (e.g., $\kappa_2 > 0$, $\kappa_2^2 \leq \kappa_4$, etc), one needs to correctly map the reduced form parameters γ into the moments of ε_{it}.

3.3.1 Alternative ways of addressing the incidental parameters problem

Instead of using jackknife, one could remove the incidental parameters bias analytically computing the bias of ψ_{it}, estimating it, and then bias-correcting the moment condition ψ_{it}. The formulas for bias correction are given in FL13, but are themselves rather involved, so the estimation of the nuisance parameters entering the bias correction terms may introduce additional finite sample bias.

4 Large Sample Theory

The framework considered in this paper is the fixed effects GMM (FE-GMM) framework of Fernández-Val and Lee (2013).

Assumption 1 (DGP)

(i) For each i, conditional on α_{i0}, $(X_i^*, S_i) \equiv \{(X_{it}^*, S_{it})\}_{t=1}^T$ is a stationary mixing sequence with strong mixing coefficients $a_i(l) = \sup_t \sup_{A \in \mathcal{A}_t^i, D \in \mathcal{D}_{t+l}^i} |\mathbb{P}(A \cap D) - \mathbb{P}(A)\mathbb{P}(D)|$, where $\mathcal{A}_t^i = \sigma(\alpha_{i0}, X_{it}^*, S_{it}, X_{it-1}^*, S_{it-1}, \dots)$ and $D_t^i = \sigma(\alpha_{i0}, X_{it}^*, S_{it}, X_{it+1}^*, S_{it+1}, \dots)$, such that $\sup_i |a_i(l)| \leqslant Ca^l$ for some $0 < a < 1$ and $C > 0$;

(ii) $\{(X_i^*, S_i, \alpha_{i0})\}_{i=1}^N$ are iid across i;

(iii) $N, T \to \infty$ such that $N/T \to \kappa^2$ for some $0 < \kappa < \infty$;

(iv) the moment function $g(\cdot)$ satisfies $\mathbb{E}[g(X_{it}^*, S_{it}, \theta_0, \alpha_{i0})] = 0$ for each i and t, where $\mathbb{E}[\cdot]$ denotes the expectation taken with respect to the distribution of (X_{it}^*, S_{it}) conditional on α_{i0}.

Assumption 1 formally specifies the data generating process and is analogous to Condition 1 in Fernández-Val and Lee (2013). Conditions (i) and (ii) require data to be independent and identically distributed across cross-sectional dimension i and impose stationarity and weak dependence over time series dimension t (conditional on the the realization of the fixed effect). Condition (iii) is a standard asymptotic approximation used to characterize the incidental parameter asymptotic bias in panels with both N and T large. Condition (iv) says that, for each i and t, the moment conditions are satisfied at the parameter of interest θ_0 and the corresponded fixed effect α_{i0}. Importantly, $\mathbb{E}[\cdot]$ denotes the expectation taken with respect to the distribution of (X_{it}^*, S_{it}) conditional on the fixed effect α_{i0} and therefore should be indexed by i or α_{i0}. For brevity, we suppress this indexing hereafter.

However, instead of $\{(X_{it}^*, S_{it})\}_{i,t=1}^{N,T}$, a researcher observes $\{(X_{it}, R_{it})\}_{i,t=1}^{N,T}$ where $X_{it} = X_{it}^* + \varepsilon_{it}$, where ε_{it} is the measurement error. As result, once evaluated at X_{it} instead of X_{it}^*, the moment conditions are no longer satisfied at θ_0, i.e. $\mathbb{E}[g(X_{it}, S_{it}, \theta_0, \alpha_{i0})] \neq 0$. Therefore, the FE-GMM estimators studied by Fernández-Val and Lee (2013) are biased since they are based on invalid moments and should be corrected. To facilitate the analysis of the asymptotic properties of the estimators we make the following assumption on the behavior of the measurement error.

Assumption 2 (MME) $\{\varepsilon_{it}\}_{i,t=1}^{N,T}$ *are iid across* i *and* t *and independent of* $\{X_{it}^*, S_{it}\}_{i,t=1}^{N,T}$ *with* $\mathbb{E}[\varepsilon_{it}] = 0$ *and* $\sigma^2 \equiv \mathbb{E}[\varepsilon_{it}^2] = o((nT)^{-1/3})$. *Suppose also that all the higher moments of* ε_{it}/σ *exist and uniformly bounded.*

On top of satisfying the classical measurement error requirements, Assumption 2 requires the variance of the measurement error to slowly shrink towards zero as the sample size grows. A similar alternative asymptotic framework was first introduced in Evdokimov and Zeleneev (2016) and proved to be a useful asymptotic approximation which suggests a simple and effective way of treatment of measurement errors in nonlinear models. Note that although the magnitude of the measurement error variance decreases with the sample size, the measurement error still affects the asymptotic distributions of the uncorrected estimators. Indeed, under Assumptions 1, 2 and weak smoothness of the moment function $g(\cdot)$, $\mathbb{E}[g(X_{it}, S_{it}, \theta_0, \alpha_{i0})] = O(\sigma^2)$. Hence, if an estimator ignores that X_{it}^* is mismeasured, it suffers from the measurement error bias of order $O(\sigma^2)$. Therefore, it will be asymptotically biased if $\sqrt{nT}\sigma^2 \to C \neq 0$ or even no longer $\sqrt{nT}$-consistent if $\sqrt{nT}\sigma^2 \to \infty$.

To account for the presence of the measurement error, as in Evdokimov and Zeleneev (2016), we introduce the corrected moment function $\psi(x, s, \theta, \alpha, \gamma) \equiv g(x, s, \theta) - \gamma g_x^{(2)}(x, s, \theta, \alpha)$, where γ is an additional nuisance parameter and $g_x^{(k)}(\cdot) \equiv \partial^k g(\cdot)/\partial x^k$. This choice is motivated by the following lemma preceded by a short notational introduction.

Let $\mathcal{X}$ be a convex set in $\mathbb{R}$ which includes the supports of X_{it}^* and X_{it} (for every sample size). Similarly, let $\mathcal{S}$ be a convex set in $\mathbb{R}^{\dim(S_{it})}$ which includes the support of S_{it}.

Lemma 1 *Suppose* $g_x^{(k)}(x, s, \theta_0, \alpha_{i0})$ *are uniformly bounded on* $\mathcal{X} \times \mathcal{S}$ *for* $k \in \{0, \ldots, 3\}$. *Then, under Assumptions 1 and 2,*

$$\mathbb{E}[\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)] = o\left((nT)^{-1/2}\right),$$

where $\gamma_0 \equiv \sigma^2/2$.

Lemma 1 says that, unlike the original moment restrictions based on $g(\cdot)$, the corrected moment restrictions based on $\psi(\cdot)$ are satisfied at the true structural parameter θ_0, the fixed effect α_{i0} for some value of the nuisance parameter $\gamma = \gamma_0$ equal to the variance of the measurement error over two.[3] Therefore, the modified moment conditions

[3]Although $\mathbb{E}[\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)]$ is not exactly equal to zero, it is sufficiently small to ensure that

are valid and can be relied on for estimation and doing asymptotically correct inference on θ_0 (provided that the incidental parameter biased is also removed).

Remark 1 *On the top of Assumptions 1 and 2, Lemma 1 requires $g(\cdot)$ to be sufficiently smooth. The boundedness requirement is for simplicity of exposition only and can be substantially weakened as in Evdokimov and Zeleneev (2016). Note that it is automatically satisfied under continuity of $g_x^{(k)}(\cdot)$ and compactness of $\mathcal{X}$ and $\mathcal{S}$ (the latter can be the case when X_{it}^*, S_{it}, and ε_{it} have bounded supports).*

In other words, once the moment function is corrected to account for the measurement error, we are back in the framework of Fernández-Val and Lee (2013) with $\psi(\cdot)$ taking place of $g(\cdot)$. Specifically, as in Fernández-Val and Lee (2013), we consider the following fixed effects GMM (FE-GMM) estimator:[4]

$$(\hat{\theta}, \{\hat{\alpha}_i\}_{i=1}^N, \hat{\gamma}) \equiv \arg \inf_{\{(\theta', \alpha_i')' \in \mathcal{B}\}_{i=1}^N, \gamma \in \Gamma} \sum_{i=1}^n \overline{\psi}_i(\theta, \alpha_i, \gamma)' \hat{W}_i^{-1} \overline{\psi}_i(\theta, \alpha_i, \gamma),$$

where $\overline{\psi}_i(\theta, \alpha_i, \gamma) \equiv T^{-1} \sum_{t=1}^T \psi(X_{it}, S_{it}, \theta, \alpha_i, \gamma)$ and $\{\hat{W}_i\}_{i=1}^N$ is a collection of weighting matrices. Here the criterion function is expressed as the sum of the individual criterion functions $\overline{\psi}_i(\theta, \alpha_i, \gamma)' \hat{W}_i^{-1} \overline{\psi}_i(\theta, \alpha_i, \gamma)$ depending on the common parameters θ and γ and the individual parameter α_i. $\mathcal{B}$ and Γ are the optimization parameter spaces for the original vector of parameters $\beta \equiv (\theta', \alpha_i')$ and the nuisance parameter γ respectively. Since the suggested estimator is based on the measurement error corrected moments, we refer to it as the moderate measurement error FE-GMM (MME-FE-GMM) estimator hereafter.

The large sample properties of the FE-GMM estimators are studied by Fernández-Val and Lee (2013). The following set of assumptions ensure that we can invoke the result of that paper to characterize the asymptotic distribution of $\hat{\theta}$.

Assumption 3 (Parameter space) *$\mathcal{B}$ and Γ are compact convex subsets of $\mathbb{R}^{d_\theta + d_\alpha}$ and $\mathbb{R}$ respectively. For each i, $\beta_{i0} \equiv (\theta_0', \alpha_{i0}')'$ is in the interior and bounded away from the boundary of $\mathcal{B}$. Similarly, (for every sample size) γ_0 is in the interior and bounded away from the boundary of Γ.*

Assumption 4 (Weighting matrices)

$(nT)^{-1/2} \sum_{i=1}^n \sum_{t=1}^T \psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0) \xrightarrow{d} N(0, \Omega^*)$ for some symmetric positive definite matrix Ω^*, so the asymptotic bias caused by the measurement error is removed.

[4] Fernández-Val and Lee (2013) call this estimator a one-step FE-GMM estimator.

(i) $\sup_{1\leqslant i\leqslant N} \left|\left|\hat{W}_i - W_i\right|\right| = o_p(1)$ *where* $\{W_i\}_{i=1}^N$ *is a deterministic sequence of symmetric positive definite matrices satisfying* $0 < 1/C < \lambda_{min}(W_i) \leqslant \lambda_{max}(W_i) < C$*;*

(ii) there exists $\xi_i(x,s)$ *uniformly bounded on* $\mathcal{X} \times S$ *such that* $\hat{W}_i = W_i + \sum_{t=1}^T \xi_i(X_{it}, S_{it})/T + R_i^W/T$*, where* $\mathbb{E}\left[\xi_i(X_{it}, S_{it})\right] = 0$ *and* $\max_i |R_i^W| = o_p(T^{1/2})$*.*

Assumption 3 is standard, it requires the parameter spaces to be compact (for consistency) and rules out the parameters on the boundary problem (for asymptotic normality). Assumption 4 governs the behavior of the weighting matrices. Condition (i) requires uniform consistency of $\{\hat{W}_i\}_{i=1}^N$ to the limiting weighting matrix sequence $\{W_i\}_{i=1}^N$ with eigenvalues uniformly bounded from below and above. Condition (ii) is a regularity condition, it is similar to Condition 4 (ii) in Fernández-Val and Lee (2013).

Remark 2 *Since the distribution of* X_{it} *drifts with the sample size,* $\xi_i(x,s)$ *also has to change to satisfy* $\mathbb{E}\left[\xi_i(X_{it}, S_{it})\right] = 0$*. To avoid this, an alternative way to formulate Condition* (ii) *is to require* $\hat{W}_i = \hat{W}_i = W_i + \sum_{t=1}^T \xi_i(X_{it}^*, S_{it})/T + R_i^W/T$ *with* $\mathbb{E}\left[\xi_i(X_{it}^*, S_{it}))\right] = 0$*, which is also not restrictive.*

Assumption 5 (Moment Functions) $\partial^{d_1+d_2} g_x^{(k)}(x,s,\theta,\alpha)/\partial\theta^{d_1}\partial\alpha^{d_2}$ *are continuous and (uniformly in i) bounded on* $\mathcal{X} \times \mathcal{S} \times \mathcal{B}$ *for* $k \in \{0, \ldots, 3\}$ *and* $0 \leqslant d_1 + d_2 \leqslant 5$*.*

Assumption 5 requires a certain degree of smoothness of the moment function and is needed for two purposes. The first purpose is the same as the one of Condition 4 (i) in Fernández-Val and Lee (2013): it governs the behavior of the higher order expansions needed to characterize the asymptotic distribution of the FE-GMM estimator. The second purpose is to localize the effect of the measurement error, i.e. bound the reminders associated with the measurement error. It helps to make sure that (a) after correction, the moment conditions are valid (Lemma 1 applies since its requirement is weaker) and (b) the limits of all needed expectations taken under the measures corresponding to the error-prone variable X_{it} (and the other variables) are equal to the expectations taken under the measure corresponding to the true X_{it}, so the asymptotic properties of the estimators depend on the true distribution of mismeasured data only.

Remark 3 *As in Remark 1, we want to point out that, in Assumption 5, boundedness is imposed for the ease of exposition and, in principle, can be replaced by weaker conditions. Instead one can require the dominance conditions as in Fernández-Val and Lee (2013) and limit the measurement error impact as in Evdokimov and Zeleneev (2016). Again, the boundedness requirement is automatically satisfied under continuity on* $\mathcal{X} \times \mathcal{S} \times \mathcal{B}$ *and compactness of* $\mathcal{X}$ *and* $\mathcal{S}$ *(we already required compactness of the parameter space).*

Similarly to Fernández-Val and Lee (2013), we introduce the following objects:

$$
\begin{aligned}
\Psi_{\theta_i} &\equiv \mathbb{E}\left[\partial\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)/\partial\theta\right] \\
\Psi_{\gamma_i} &\equiv \mathbb{E}\left[\partial\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)/\partial\gamma\right] = -\mathbb{E}\left[g_x^{(2)}(X_{it}, S_{it}, \theta_0, \alpha_{i0})\right], \\
\Psi_{(\theta_i,\gamma_i)} &\equiv [\Psi_{\theta_i}, \Psi_{\gamma_i}], \\
\Psi_{\alpha_i} &\equiv \mathbb{E}\left[\partial\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)/\partial\alpha\right], \\
P_{\alpha_i} &\equiv W_i^{-1} - W_i^{-1}\Psi_{\alpha_i}(\Psi'_{\alpha_i}W_i^{-1}\Psi_{\alpha_i})^{-1}\Psi'_{\alpha_i}W_i^{-1}, \\
J &\equiv \overline{\mathbb{E}}\left[\Psi'_{(\theta_i,\gamma_i)}P_{\alpha_i}\Psi_{(\theta_i,\gamma_i)}\right],
\end{aligned}
$$

where $\overline{\mathbb{E}}[\cdot]$ denotes the expectation taken with respect to the distribution of α_{i0}. Finally, let Ω_i denote the (conditional) long run variance of $\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)$. Specifically,

$$
\begin{aligned}
\Omega_i &= \Omega_{0i} + \sum_{j=1}^{\infty}(\Omega_{ji} + \Omega'_{ji}), \\
\Omega_{ji} &= \mathbb{E}\left[\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)\psi(X_{it-j}, S_{it-j}, \theta_0, \alpha_{i0}, \gamma_0)'\right].
\end{aligned}
$$

Assumption 6 (ID and regularity)

(i) For each $\eta > 0$,

$$
\liminf_{N,T\to\infty} \inf_i \left[\inf_{(\theta',\alpha_i')'\in\mathcal{B},\gamma\in\Gamma:||(\theta',\alpha_i',\gamma)'-(\theta_0',\alpha_{i0}',\gamma_0)'||>\eta} ||\psi_i(\theta, \alpha_i, \gamma)|| \right] > 0,
$$

where $\psi_i(\theta, \alpha_i, \gamma) \equiv \mathbb{E}\left[\psi(X_{it}, S_{it}, \theta, \alpha_i, \gamma)\right]$;

(ii) $\lambda_{min}(\Psi'_{\alpha_i}W_i^{-1}\Psi_{\alpha_i}) > C > 0$ uniformly in i and $\lambda_{min}(J) > C > 0$;

(iii) $0 < 1/C < \lambda_{min}(\Omega_i) \leqslant \lambda_{max}(\Omega_i) < C$ uniformly in i.

Assumption 6 is a collection identification and regularity conditions. Condition (i) is a global identification condition. Note that, unlike in Fernández-Val and Lee (2013), the distribution of (X_{it}, S_{it}) necessarily drifts and we need to use $\liminf_{N,T\to\infty}$ in from of $\inf_i$ (Remark 4 below also addresses this point). Like Condition 2 (v) in Fernández-Val and Lee (2013), it is needed to be uniform to establish uniform consistency of the estimators of the individual parameters. Condition (ii) is a local identification condition which allows to establish asymptotic normality of the estimators: the first part correspond to the individual parameters $\{\hat{\alpha}_i\}_{i=1}^N$ and the second to the common parameters $\hat{\theta}$ and $\hat{\gamma}$.

Condition (iii) is a standard regularity condition imposed on the long run variance of the moment conditions.

Remark 4 *For simplicity of exposition, we stated Assumption 6 using objects of the form* $f_i \equiv \mathbb{E}\left[f(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)\right]$ *for generic function* $f(\cdot)$*. These expectations change with the sample size because both the distribution of* X_{it} *and the value of* $\gamma_0 = \sigma^2/2$ *drift. However, note that, under the moderate measurement error asymptotics, for a smooth function* $f(x, s, \theta, \alpha, \gamma)$*,* $f_i \to f_i^* \equiv \mathbb{E}\left[f(X_{it}^*, S_{it}, \theta_0, \alpha_{i0}, 0)\right]$ *as* $N, T \to \infty$*. In other words, once evaluating the expectations,* X_{it} *can be replaced by* X_{it}^* *and* γ_0 *can be replaced by 0 in the limit. Note that the limiting "star" object* f_i^* *does not depend on the features of the measurement error distribution at all and is determined by the underlying statistical model for* (X_{it}^*, S_{it}) *only. As a result, Assumption 6 can be formulated using the limiting "star" objects only. For example, in Condition* (i) $\psi_i(\theta, \alpha_i, \gamma)$ *can be replaced by* $\psi_i^*(\theta, \alpha_i, \gamma) \equiv \mathbb{E}\left[\psi(X_{it}^*, S_{it}, \theta, \alpha_i, \gamma)\right]$ *(and, consequently,* $\liminf_{N,T\to\infty}$ *can be dropped). In condition* (ii)*,* Ψ_{α_i} *and* J *can be replaced by* $\Psi_{\alpha_i}^* \equiv \mathbb{E}\left[\partial\psi(X_{it}^*, S_{it}, \theta_0, \alpha_{i0}, 0)/\partial\alpha\right]$ *and* $J^* \equiv \overline{\mathbb{E}}[\Psi^{*\prime}_{(\theta_i,\gamma_i)} P_{\alpha_i}^* \Psi^*_{(\theta_i,\gamma_i)}]$*, respectively, with* $\Psi^*_{(\theta_i,\gamma_i)}$ *and* $P_{\alpha_i}^*$ *defined in the same way as the other "star" objects before. Similarly, since* $\psi(x, s, \theta, \alpha, 0) = g(x, s, \theta, \alpha)$*, in Condition* (iii) Ω_i *can be replaced by* Ω_i^**, the long run variance of the original moments* $g(X_{it}^*, S_{it}, \theta_0, \alpha_{i0})$*. Therefore, Assumption 6 is, in fact, not related to the features of the measurement error distribution at all: it can be equivalently formulated and verified in terms of the original statistical model for* (X_{it}^*, S_{it}) *only.*

Essentially, Assumptions 1-6 allow to establish validity of the corrected moments and to verify that Conditions 1-4 of Fernández-Val and Lee (2013) are satisfied with $\psi(\cdot)$ taking place of the original moment function $g(\cdot)$. Therefore, applying Theorem 2 in Fernández-Val and Lee (2013), we can characterize the asymptotic distribution of the estimator of the common parameters $\hat{\zeta} \equiv (\hat{\theta}', \hat{\gamma}')'$:

Theorem 1 *Under Assumptions 1-6,*

$$(nT)^{1/2}\Xi^{-1/2}(\hat{\zeta} - \zeta_0 - B/T) \xrightarrow{d} N(0, I_{d_\theta+1}), \tag{19}$$

where $\zeta_0 \equiv (\theta_0', \gamma_0')'$*,* $\Xi \equiv J^{-1}VJ^{-1}$*,* $V \equiv \overline{\mathbb{E}}\left[\Psi'_{(\theta_i,\gamma_i)} P_{\alpha_i} \Omega_i P_{\alpha_i} \Psi_{(\theta_i,\gamma_i)}\right]$*, and the expression for* B *is given in the Appendix.*

According to Theorem 1, the estimator of the common parameters $\hat{\zeta}$ suffers from the bias of the order $1/T$ which is typical in the large N and T panel literature. Since the estimator is based on the corrected moments, the measurement error part of the bias does not appear in the asymptotic distribution, and the only source of the bias left is the incidental parameter bias B/T. As usual, under $N/T \to \kappa^2$ asymptotics, the asymptotic approximation of the distribution of $(nT)^{1/2}(\hat{\zeta} - \zeta_0)$ is $N(\kappa B, \Xi)$, so the estimator is asymptotically biased by κB and needed to be corrected in order to be a base for asymptotically valid inference.

Remark 5 *To characterize the distribution of the individual effects estimators* $\{\hat{\alpha}_i\}_{i=1}^{N}$*, one can apply the result of Lemma 1 in Fernández-Val and Lee (2013) which requirements are also satisfied under Assumptions 1-6.*

Remark 6 *As pointed in Remark 4, under the moderate measurement error asymptotics, the objects, which control the asymptotic distribution of* $\hat{\zeta}$*, like* Ξ *and* B *can be replaced their their limits, analogous "star" objects. Then, similarly to Assumption 6, the asymptotic distribution of* $\hat{\zeta}$ *in fact does not depend on the features of the measurement error distribution and is controlled by the underlying statistical model for the correctly measured data* (X_{it}^{*}, S_{it}) *only. Therefore* (19) *can be alternatively represented as*

$$(nT)^{1/2}\Xi^{*-1/2}(\hat{\zeta} - \zeta_0 - B^{*}/T) \xrightarrow{d} N(0, I_{d_\theta+1}),$$

where $\Xi^{*} \equiv J^{*-1}V^{*}J^{*-1}$ *and* $V^{*} \equiv \overline{\mathbb{E}}\left[\Psi^{*\prime}_{(\theta_i,\gamma_i)} P^{*}_{\alpha_i}\Omega^{*}_i P^{*}_{\alpha_i}\Psi^{*}_{(\theta_i,\gamma_i)}\right]$*. The expression for* B^{*} *is also given in the appendix.*

In order to make inference on the structural parameter θ_0 (or a function of θ_0) based on the MME-FE-GMM estimator $\hat{\zeta}$, one just needs to correct for the incidental parameter bias B/T (for example, by jackknifing or analytically) and estimate the asymptotic variance of the corrected estimator. In many large N and T panel settings, removing the incidental parameter bias does not increase the asymptotic variance of the estimator. For example, Fernández-Val and Lee (2013) propose three analytical bias correction methods, which do not affect it. However, since dealing with the incidental parameter bias is not the focus of the paper, we do not propose and study properties of any specific bias reduction technique, but just provide an estimator of the asymptotic variance Ξ in (19), which, as pointed before, in many standard settings is unaffected by the standard incidental parameter bias correction methods. Specifically, one can compute $\hat{\Xi}$ using the following formulas:

$$\begin{aligned}
\hat{\Psi}_{\theta_i} &\equiv T^{-1}\sum_{t=1}^{T}\partial\psi(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i,\hat{\gamma})/\partial\theta \\
\hat{\Psi}_{\gamma_i} &\equiv T^{-1}\sum_{t=1}^{T}\partial\psi(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i,\hat{\gamma})/\partial\gamma = -T^{-1}\sum_{t=1}^{T}g_x^{(2)}(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i), \\
\hat{\Psi}_{(\theta_i,\gamma_i)} &\equiv [\hat{\Psi}_{\theta_i},\hat{\Psi}_{\gamma_i}], \\
\hat{\Psi}_{\alpha_i} &\equiv T^{-1}\sum_{t=1}^{T}\partial\psi(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i,\hat{\gamma})/\partial\alpha, \\
\hat{P}_{\alpha_i} &\equiv W_i^{-1} - W_i^{-1}\hat{\Psi}_{\alpha_i}(\hat{\Psi}'_{\alpha_i}W_i^{-1}\hat{\Psi}_{\alpha_i})^{-1}\hat{\Psi}'_{\alpha_i}W_i^{-1}, \\
\hat{J} &\equiv N^{-1}\sum_{i=1}^{N}\hat{\Psi}'_{(\theta_i,\gamma_i)}\hat{P}_{\alpha_i}\hat{\Psi}_{(\theta_i,\gamma_i)}, \\
\hat{V} &\equiv N^{-1}\sum_{i=1}^{N}\hat{\Psi}'_{(\theta_i,\gamma_i)}\hat{P}_{\alpha_i}\hat{\Omega}_i\hat{P}_{\alpha_i}\hat{\Psi}_{(\theta_i,\gamma_i)}, \\
\hat{\Xi} &\equiv \hat{J}^{-1}\hat{V}\hat{J}^{-1},
\end{aligned}$$

where $\hat{\Omega}_i$ estimates Ω_i, the long run variance of $\psi(X_{it},S_{it},\theta_0,\alpha_{i0},\gamma_0)$.

Remark 7 *The moderate measurement error asymptotic provides an alternative way of estimating Ξ, which exploits the knowledge that $\gamma_0 = \sigma^2/2 \to 0$. This implies that, in the formulas provided above, $\hat{\gamma}$ can be replaced by 0. Specifically, $\hat{\Psi}_{\theta_0}$ and $\hat{\Psi}_{\alpha_i}$ can be replaced by*

$$\begin{aligned}
\hat{\Psi}^0_{\theta_i} &\equiv T^{-1}\sum_{t=1}^{T}\partial\psi(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i,0)/\partial\theta = T^{-1}\sum_{t=1}^{T}\partial g(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i)/\partial\theta, \\
\hat{\Psi}^0_{\alpha_i} &\equiv T^{-1}\sum_{t=1}^{T}\partial\psi(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i,0)/\partial\alpha = T^{-1}\sum_{t=1}^{T}\partial g(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i)/\partial\alpha,
\end{aligned}$$

respectively. Moreover, $\hat{\Omega}_i$ can be replaced by $\hat{\Omega}_0$, which estimates the long-run variance of $\psi(X_{it},S_{it},\theta_0,\alpha_{i0},0) = g(X_{it},S_{it},\theta_0,\alpha_{i0})$ (which is, as pointed in Remark 4, is equal to the long run variance of $g(X^*_{it},S_{it},\theta_0,\alpha_{i0})$ in the limit). In particular, if $g(X^*_{it},S_{it},\theta_0,\alpha_{i0})$ is a martingale difference sequence, then $\Omega^*_i = \mathbb{E}\left[g(X^*_{it},S_{it},\theta_0,\alpha_{i0})g(X^*_{it},S_{it},\theta_0,\alpha_{i0})'\right]$, and one can estimate it by

$$\hat{\Omega}^0_i = T^{-1}\sum_{t=1}^{T}g(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i)g(X_{it},S_{it},\hat{\theta},\hat{\alpha}_i)'.$$

5 Large Network Data Models

[...]

References

Amemiya, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *Journal of Econometrics*, 28(3):273–289.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. In Blundell, R., Newey, W., and Persson, T., editors, *Advances in Economics and Econometrics*, pages 381–409. Cambridge University Press.

Arellano, M. and Honoré, B. (2001). Chapter 53 - panel data models: Some recent developments. In Heckman, J. J. and Leamer, E., editors, *Handbook of Econometrics*, volume 5, pages 3229 – 3296. Elsevier.

Bao, Y. and Ullah, A. (2007). The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics*, 140(2):650–669.

Chen, M., Fernández-Val, I., and Weidner, M. (2014). Nonlinear factor models for network and panel data. *Working Paper*.

Chen, X., Hong, H., and Nekipelov, D. (2011). Nonlinear models of measurement errors. *Journal of Economic Literature*, 49(4):901–937.

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

Chesher, A. (1991). The effect of measurement error. *Biometrika*, 78(3):451–462.

Dhaene, G. and Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies*, 82(3):991–1030.

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Princeton University.

Evdokimov, K. S. and Zeleneev, A. (2016). Simple estimation of semiparametric models with measurement errors. *Working Paper, Princeton University.*

Fernández-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*, 150(1):71–85.

Fernandez-Val, I. and Lee, J. (2012). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *arXiv:1206.2966 [econ, math, stat]*. arXiv: 1206.2966.

Fernández-Val, I. and Lee, J. (2013). Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics*, 4(3):453–481.

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1):93–118.

Hahn, J. and Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 27(06):1152–1191.

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Hausman, J. A., Ichimura, H., Newey, W. K., and Powell, J. L. (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics*, 50(3):273–295.

Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611–629.

Linton, O. and Shiu, J.-L. (2018). Semiparametric nonlinear panel data models with measurement error. CeMMAP working papers CWP09/18, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica*, 72(1):219–255.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Rilstone, P., Srivastava, V. K., and Ullah, A. (1996). The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics*, 75(2):369–395.

Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica*, 75(1):201–239.

Schennach, S. M. (2013). Measurement error in nonlinear models – a review. In Acemoglu, D., Arellano, M., and Dekel, E., editors, *Advances in Economics and Econometrics*, pages 296–337.

Schennach, S. M. (2016). Recent advances in the measurement error literature. *Annual Review of Economics*, 8(1):341–377.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Wansbeek, T. (2001). GMM estimation in panel data models with measurement error. *Journal of Econometrics*, 104(2):259–268.

Wilhelm, D. (2015). Identification and estimation of nonparametric panel data regressions with measurement error. CeMMAP working papers CWP34/15.

Wolter, K. M. and Fuller, W. A. (1982). Estimation of nonlinear errors-in-variables models. *The Annals of Statistics*, 10(2):539–548.

A Proofs

A.1 Proofs of Propositions in Section 2

A.1.1 Proof of Proposition 4

1. For any θ let

$$\widehat{\alpha}_i(\theta) \equiv \underset{\alpha_i}{\operatorname{argmax}} \sum_{t=1}^{T} \ell_{it}(\theta, \alpha_i).$$

Then $\widehat{\theta}_\ell$ solves

$$\begin{aligned} 0 &= E_{NT}\left[u_{it}\left(\widehat{\theta}_\ell, \widehat{\alpha}_i\left(\widehat{\theta}_\ell\right)\right)\right] \\ &= E_{NT}\left[u_{it}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right)\right] + E_{NT}\left[u_{it}^{\theta}\left(\tilde{\theta}, \widehat{\alpha}_i\left(\tilde{\theta}\right)\right) + u_{it}^{\alpha}\left(\tilde{\theta}, \widehat{\alpha}_i\left(\tilde{\theta}\right)\right) \cdot \nabla_\theta \widehat{\alpha}_i\left(\tilde{\theta}\right)\right]\left(\widehat{\theta}_\ell - \theta_0\right), \end{aligned}$$

and hence

$$\sqrt{NT}\left(\widehat{\theta}_\ell - \theta_0\right) = -E_{NT}\left[u_{it}^{\theta}\left(\tilde{\theta}, \widehat{\alpha}_i\left(\tilde{\theta}\right)\right) + u_{it}^{\alpha}\left(\tilde{\theta}, \widehat{\alpha}_i\left(\tilde{\theta}\right)\right) \cdot \nabla_\theta \widehat{\alpha}_i\left(\tilde{\theta}\right)\right]^{-1} \sqrt{NT} E_{NT}\left[u_{it}\left(\theta_0, \widehat{\alpha}_i(\theta_0)\right)\right] \tag{A.1}$$

2. Consider $E_{NT}\left[u_{it}\left(\theta_0, \widehat{\alpha}_i\left(\theta_0\right)\right)\right]$. Since $\widehat{\alpha}_i\left(\theta_0\right)$ solves

$$E_T\left[v_{it}\left(\theta_0, \widehat{\alpha}_i\left(\theta_0\right)\right)\right] = 0,$$

we have

$$\begin{aligned} E_{NT}\left[u_{it}\left(\theta_0, \widehat{\alpha}_i\left(\theta_0\right)\right)\right] &= E_{NT}\left[u_{it}\left(\theta_0, \alpha_{i0}\right)\right] + E_N\left[A_{ni}\right] + R_{n3}, \\ A_{ni} &\equiv \frac{1}{T}\sum_{t=1}^{T}\left\{u_{it}^{\alpha}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right) + \frac{1}{2}u_{it}^{\alpha}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right)^2\right\} \quad \text{(A.2)} \\ \left|R_{n3}\right| &\leq CE_N\left[\left|\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right|^3\right]. \end{aligned}$$

Because of the measurement error, $E\left[u_{it}\left(\theta_0, \alpha_{i0}\right)\right] \neq 0$, and

$$\begin{aligned} E_{NT}\left[u_{it}\right] &= E_{NT}\left[u_{it}^*\right] + E_{NT}\left[u_{itx}^*\varepsilon_{it}\right] + \frac{1}{2}E_{NT}\left[u_{itxx}^*\varepsilon_{it}^2\right] + O_P\left(\sigma^3\right) \\ &= E_{NT}\left[u_{it}^*\right] + \frac{1}{2}\sigma^2 E\left[u_{itxx}^*\right] + E_{NT}\left[u_{itx}^*\varepsilon_{it}\right] + o_P\left(T^{-1-3\delta}\right). \end{aligned}$$

Moreover, we cannot make use of the existing results on the higher order expansions of m-estimators, such as Rilstone et al. (1996) and Bao and Ullah (2007), because those expansions are invalidated by the presence of measurement error. Instead, in Lemma **??** below, we show that

$$\widehat{\alpha}_i - \alpha_0 = E_T\left[\psi_{it}^*\right] + \sigma^2\beta_{\mathrm{ME},i} + \frac{1}{T}\beta_{\mathrm{INC},i} + E_T\left[\zeta_{it}\right] + o_P\left(T^{-1-3\delta}\right),$$

where $E\zeta_{it} = 0$, $E\zeta_{it}^2 = O\left(\sigma^2\right)$, $E_T\left[\zeta_{it}\right] = O_P\left(T^{-1/2}\sigma\right) = o_P\left(T^{-5/6-\delta}\right)$. Hence,

$$\begin{aligned} A_{ni} &\equiv E_T\left[u_{it}^{\alpha}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right) + \frac{1}{2}u_{it}^{\alpha\alpha}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right)^2\right] \\ &= E_T\left[u_{it}^{*\alpha}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right) + \frac{1}{2}u_{it}^{*\alpha\alpha}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right)^2\right] \\ &\quad + E_T\left[u_{itx}^{*\alpha}\varepsilon_{it}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right) + \frac{1}{2}u_{itx}^{*\alpha\alpha}\varepsilon_{it}\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right)^2\right] \\ &\quad + \frac{1}{2}E_T\left[u_{itxx}^{*\alpha}\varepsilon_{it}^2\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right) + \frac{1}{2}u_{itxx}^{*\alpha\alpha}\varepsilon_{it}^2\left(\widehat{\alpha}_i\left(\theta_0\right) - \alpha_{i0}\right)^2\right] + o_P\left(T^{-1-3\delta}\right) \\ &\quad + E_T\left[\rho_{it}\right] + R_{ni}^A \\ &= E_i\left[u_{it}^{*\alpha}\right]E_T\left[\psi_{it}^*\right] + \frac{1}{T}\mathbb{C}_{\mathrm{LR},i}\left[u_{it}^{*\alpha}, \psi_{it}^*\right] + E_i\left[u_{it}^{*\alpha}\right]\left(\frac{1}{T}\beta_{\mathrm{INC},i} + \sigma^2\beta_{\mathrm{ME},i}\right) \\ &\quad + \frac{1}{2T}E_i\left[u_{it}^{*\alpha\alpha}\right]V_{\mathrm{LR},i}\left[\psi_{it}^*\right] + o_P\left(T^{-1-3\delta}\right) + E_T\left[\rho_{it}\right] + R_{ni}^A, \end{aligned}$$

where $\rho_{it} \equiv \rho\left(X_{it}^*, \varepsilon_{it}, S_{it}\right)$ for some bounded function ρ, and $E\rho_{it} = 0$, $E\rho_{it}^2 = o\left(1\right)$.

Suppose $\max_{i\in[N]}|R^A_{ni}| = o_P(T^{-1})$ and $\max_{i\in[N]}|\widehat{\alpha}_i(\theta_0)-\alpha_{i0}|^3 = o_P(T^{-1})$. Then,

$$\sqrt{NT}\left(E_{NT}u_{it}(\theta_0,\widehat{\alpha}_i(\theta_0)) - \mathcal{B}^u_{T,\sigma} + o\left(T^{-1-3\delta}\right)\right) \to_d N\left(0, V_{\text{LR},}\left[u^*_{it} + E_i\left[u^{*\alpha}_{it}\right]\psi^*_{it}\right]\right), \tag{A.3}$$

where

$$\mathcal{B}^u_{T,\sigma} \equiv \sigma^2 E\left[\frac{1}{2}u^*_{itxx} + u^{*\alpha}_{it}\beta_{\text{ME},i}\right] + \frac{1}{T}E\left[u^{*\alpha}_{it}\beta_{\text{INC},i} + \mathbb{C}_{\text{LR},i}\left[u^{*\alpha}_{it},\psi^*_{it}\right] + \frac{1}{2}u^{*\alpha\alpha}_{it}V_{\text{LR},i}\left[\psi^*_{it}\right]\right]. \tag{A.4}$$

3. Consider the first term on the right-hand side of equation (A.1). Since $\widehat{\alpha}_i(\theta)$ solves $E_{NT}\left[v_{it}(\theta,\widehat{\alpha}_i(\theta))\right] = 0$, by the implicit function theorem,

$$\nabla_\theta\widehat{\alpha}_i(\theta) = -E_T\left[v^\alpha_{it}(\theta,\widehat{\alpha}_i(\theta))\right]^{-1}E_T\left[v^\theta_{it}(\theta,\widehat{\alpha}_i(\theta))\right].$$

Let $H^* \equiv -\left(E\left[u^{*\theta}_{it} - u^{*\alpha}_{it}\cdot E_i\left[v^{*\theta}_{it}\right]/E_i\left[v^{*\alpha}_{it}\right]\right]\right)^{-1}$,Then $\sqrt{NT}\left(\widehat{\theta}_\ell - \theta_0\right) = H^*\sqrt{NT}E_{NT}\left[u_{it}(\theta_0,\widehat{\alpha}_i(\theta_0))\right] + o_P(1)$,and hence

$$\sqrt{NT}\left(\widehat{\theta}_\ell - \theta_0 - H^*\mathcal{B}^u_{T,\sigma}\right) \to_d N\left(0, H^*V_{\text{LR},}\left[u^*_{it} + E_i\left[u^{*\alpha}_{it}\right]\psi^*_{it}\right]H^*\right).$$

4. To establish the above result formally one needs to show that $\max_{i\in[N]}|R^A_{ni}| = o_P(T^{-1})$ and $\max_{i\in[N]}|\widehat{\alpha}_i(\theta_0)-\alpha_{i0}|^3 = o_P(T^{-1})$. We do not do this in this for the biased m-estimators in this proposition. We do formally establish asymptotic normality of the MME estimator in Section 3 that we recommend researchers use in practice. ■

A.2 Proofs of Large Sample Theory Results

Proof. [Proof of Lemma 1] By expanding $g(X_{it},S_{it},\theta_0,\alpha_{i0})$ around X^*_{it} and using $X_{it} = X^*_{it} + \varepsilon_i$, we get

$$\begin{aligned} g(X_{it},S_{it},\theta_0,\alpha_{i0}) =& g(X^*_{it},S_{it},\theta_0,\alpha_{i0}) + g^{(1)}_x(X^*_{it},S_{it},\theta_0,\alpha_{i0})\varepsilon_{it} \\ &+ \frac{1}{2}g^{(2)}_x(X^*_{it},S_{it},\theta_0,\alpha_{i0})\varepsilon^2_{it} + \frac{1}{6}g^{(3)}_x(\tilde{X}_{it},S_{it},\theta_0,\alpha_{i0})\varepsilon^3_{it}, \end{aligned}$$

where X_{it} lies between X^*_{it} and X_{it}. Note that, by Assumption 2 and boundedness of $g^{(3)}_x(x,s,\theta_0,\alpha_{i0})$, $\mathbb{E}\left[\frac{1}{6}g^{(3)}_x(\tilde{X}_{it},S_{it},\theta_0,\alpha_{i0})\varepsilon^3_{it}\right]$ exists and is of order $\mathbb{E}\left[|\varepsilon_{it}|^3\right] = O(\sigma^3) = o((nT)^{-1/2})$. Therefore, using independence of ε_{it} of (X^*_{it},S_{it}),

$$\mathbb{E}\left[g(X_{it},S_{it},\theta_0,\alpha_{i0})\right] = \frac{\sigma^2}{2}\mathbb{E}\left[g^{(2)}_x(X^*_{it},S_{it},\theta_0,\alpha_{i0})\right] + o((nT)^{-1/2}).$$

Similarly,

$$\mathbb{E}\left[g_x^{(2)}(X_{it}, S_{it}, \theta_0, \alpha_{i0})\right] = \mathbb{E}\left[g_x^{(2)}(X_{it}^*, S_{it}, \theta_0, \alpha_{i0})\right] + O(\sigma),$$

and, as a result,

$$\begin{aligned}\mathbb{E}\left[\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)\right] &= \mathbb{E}\left[g(X_{it}, S_{it}, \theta_0, \alpha_{i0})\right] - \frac{\sigma^2}{2}\mathbb{E}\left[g_x^{(2)}(X_{it}, S_{it}, \theta_0, \alpha_{i0})\right] \\ &= o((nT)^{-1/2}) + O(\sigma^3) \\ &= o((nT)^{-1/2}),\end{aligned}$$

where we exploited $O(\sigma^3) = o((nT)^{-1/2})$ again. ■

Proof. [Proof of Theorem 1] The result of the theorem is obtained by an application of Theorem 2 in Fernández-Val and Lee (2013) with $\psi(\cdot)$ in place of $g(\cdot)$. Therefore, we need to verify that Conditions 1-4 of Fernández-Val and Lee (2013) are satisfied for $\psi(\cdot)$ under Assumptions 1-6. Before proceeding with that, first, note that Lemma 1 establishes validity of the corrected moment condition.

Verifying Condition 1: Assumptions 1 and 2 guarantee that Condition 1 is satisfied. Indeed, the parts (ii)-(iv) of Condition 1 are trivially satisfied and we just need to verify that part (i), the mixing condition, is also satisfied. This follows from Assumption 1 (i) and 2.

Verifying Condition 2: Parts (i) and (iv) of Condition 2 are ensured by Assumption 5. Part (ii) follows from Assumption 3. Part (iii) is trivial. The first part of (v) follows directly from Assumptions 4 (i). We also need to show that for each $\eta > 0$

$$\liminf_{N,T\to\infty} \inf_i \left[Q_i^W(\theta_0, \alpha_{i0}, \gamma_{i0}) - \inf_{(\theta',\alpha_i')'\in\mathcal{B},\gamma\in\Gamma:||(\theta',\alpha_i',\gamma)'-(\theta_0',\alpha_{i0}',\gamma_0)'||>\eta} Q_i^W(\theta, \alpha, \gamma)\right] > 0, \quad \text{(A.5)}$$

where $Q_i^W(\theta, \alpha, \gamma) \equiv -\psi_i(\theta, \alpha, \gamma)' W_i \psi_i(\theta, \alpha, \gamma)$. To prove that it is enough to show

$$\liminf_{N,T\to\infty} \inf_i Q_i^W(\theta_0, \alpha_{i0}, \gamma_{i0}) - \limsup_{N,T\to\infty} \sup_i \inf_{(\theta',\alpha_i')'\in\mathcal{B},\gamma\in\Gamma:||(\theta',\alpha_i',\gamma)'-(\theta_0',\alpha_{i0}',\gamma_0)'||>\eta} Q_i^W(\theta, \alpha, \gamma) > 0. \quad \text{(A.6)}$$

Replicating the proof of Lemma 1 and making use of the boundedness of the $g_x^{(3)}(x, s, \theta_0, \alpha_{i0})$ one can show that

$$\sup_i \psi_i(\theta_0, \alpha_{i0}, \gamma_0) = \sup_i \mathbb{E}\left[\psi(X_{it}, S_{it}, \theta_0, \alpha_{i0}, \gamma_0)\right] = o((nT)^{-1/2}),$$

Hence, using Assumption 4 (i), we conclude that $\liminf_{N,T\to\infty} \inf_i Q_i^W(\theta_0, \alpha_{i0}, \gamma_0) = 0$.

At the same time, using Assumptions 6 (i) and 4 (i), we obtain that for each $\eta > 0$

$$\limsup_{N,T\to\infty} \sup_i \inf_{(\theta',\alpha_i')'\in\mathcal{B},\gamma\in\Gamma:||(\theta',\alpha_i',\gamma)'-(\theta_0',\alpha_{i0}',\gamma_0)'||>\eta} Q_i^W(\theta,\alpha,\gamma) < 0.$$

Therefore, we conclude that (A.6) and, consequently, (A.5) both hold. Note that, unlike in Fernández-Val and Lee (2013), since the distribution of (X_{it}, S_{it}) drifts, we have $\liminf_{N,T\to\infty}$ in front of $\inf_i$.

Verifying Condition 3: Part (i) is implied by Assumption 3. Part (ii) follows from Assumption 6 (ii).

Verifying Condition 4: Part (i) is implied by Assumption 5. Part (ii) follows from Assumption 4 (ii). ■