Quantitative Economics 4 (2013), 453–481 1759-7331/20130453

Panel data models with nonadditive unobserved heterogeneity: Estimation and inference

Iván Fernández-Val
Department of Economics, Boston University

Joonhwah Lee
Department of Economics, MIT

This paper considers fixed effects estimation and inference in linear and nonlinear panel data models with random coefficients and endogenous regressors. The quantities of interest—means, variances, and other moments of the random coefficients—are estimated by cross sectional sample moments of generalized method of moments (GMM) estimators applied separately to the time series of each individual. To deal with the incidental parameter problem introduced by the noise of the within-individual estimators in short panels, we develop bias corrections. These corrections are based on higher-order asymptotic expansions of the GMM estimators and produce improved point and interval estimates in moderately long panels. Under asymptotic sequences where the cross sectional and time series dimensions of the panel pass to infinity at the same rate, the uncorrected estimators have asymptotic biases of the same order as their asymptotic standard deviations. The bias corrections remove the bias without increasing variance. An empirical example on cigarette demand based on Becker, Grossman, and Murphy (1994) shows significant heterogeneity in the price effect across U.S. states.

Keywords. Correlated random-coefficient model, panel data, instrumental variables, GMM, fixed effects, bias, incidental parameter problem, cigarette demand.

JEL classification. C23, J31, J51.

1. Introduction

This paper considers estimation and inference in linear and nonlinear panel data models with random coefficients and endogenous regressors. The quantities of interest are means, variances, and other moments of the distribution of the random coefficients. In a state level panel model of rational addiction, for example, we might be interested in

Iván Fernández-Val: ivanf@bu.edu
Joonhwah Lee: jhlee82@mit.edu

This paper is based in part on the second chapter of Fernández-Val (2005). We wish to thank Josh Angrist, Victor Chernozhukov, and Whitney Newey for encouragement and advice. For suggestions and comments, we are grateful to Manuel Arellano, Mingli Chen, the editor, three anonymous referees, and the participants at the Brown and Harvard–MIT Econometrics seminar. We thank Aju Fenn for providing the data for the empirical example. All remaining errors are ours. Fernández-Val gratefully acknowledges financial support from Fundación Caja Madrid, Fundación Ramón Areces, and the National Science Foundation.

DOI: 10.3982/QE75

the mean and variance of the distribution of the price effect on cigarette consumption across states, controlling for endogenous past and future consumption. These models pose important challenges in estimation and inference if the relation between the regressors and random coefficients is left unrestricted. Fixed effects methods based on generalized method of moments (GMM) estimators applied separately to the time series of each individual can be severely biased due to the incidental parameter problem. The source of the bias is the finite-sample bias of GMM if some of the regressors are endogenous or the model is nonlinear in parameters, or nonlinearities if the parameter of interest is the variance or other high-order moment of the random coefficients. Neglecting the heterogeneity and imposing fixed coefficients does not solve the problem, because the resulting estimators are generally inconsistent for the mean of the random coefficients (Yitzhaki (1996), and Angrist, Graddy, and Imbens (2000)).[1] Moreover, imposing fixed coefficients does not allow us to estimate other moments of the distribution of the random coefficients.

We introduce a class of bias-corrected panel fixed effects GMM estimators. Thus, instead of imposing fixed coefficients, we estimate different coefficients for each individual using the time series observations and we correct the sample moments of the estimated coefficients for the incidental parameter bias. For linear models, in addition to the bias correction, these estimators differ from the standard fixed effects estimators in that both the intercept and the slopes are different for each individual. Moreover, unlike for the classical random-coefficient estimators, they do not rely on any restriction in the relationship between the regressors and random coefficients; see Hsiao and Pesaran (2004) for a recent survey on random coefficient models. This flexibility allows us to account for Roy-type (Roy (1951)) selection where the regressors are decision variables with levels determined by their returns. Linear models with Roy selection are commonly referred to as correlated random-coefficient models in the panel data literature. In the presence of endogenous regressors, treating the random coefficients as fixed effects is also convenient to overcome the identification problems in these models pointed out by Kelejian (1974).

The most general models we consider are semiparametric in the sense that the distribution of the random coefficients is unspecified and the parameters are identified from moment conditions. These conditions can be nonlinear functions in parameters and variables, accommodating both linear and nonlinear random-coefficient models, and allowing for the presence of time varying endogeneity in the regressors that is not captured by the random coefficients. We use the moment conditions to estimate the model parameters and other quantities of interest via GMM methods applied separately to the time series of each individual. The resulting estimates can be severely biased in short panels due to the incidental parameters problem, which in this case is a consequence of the finite-sample bias of GMM (Newey and Smith (2004)) and/or the nonlinearity of the quantities of interest with respect to the random coefficients. We develop analytical corrections to reduce the bias.

[1] Heckman and Vytlacil (2000) and Angrist (2004) found sufficient conditions for fixed-coefficient ordinary least squares (OLS) and instrumental variable (IV) estimators to be consistent for the mean coefficient.

To derive the bias corrections, we use higher-order expansions of the GMM estimators, extending the analysis in Newey and Smith (2004) for cross sectional estimators to panel data estimators with fixed effects and serial dependence. If n and T denote the cross sectional and time series dimensions of the panel, the corrections remove the leading term of the bias of order $O(T^{-1})$, and center the asymptotic distribution at the true parameter value under sequences where n and T grow at the same rate. This approach is aimed at performing well in econometric applications that use moderately long panels, where the most important part of the bias is captured by the first term of the expansion. Other previous studies that used a similar approach for the analysis of linear and nonlinear fixed effects estimators in panel data include, among others, Kiviet (1995), Phillips and Moon (1999), Alvarez and Arellano (2003), Hahn and Kuersteiner (2002, 2011), Lancaster (2002), Woutersen (2002), and Hahn and Newey (2004). See Arellano and Hahn (2007) for a survey of this literature and additional references.

The first distinctive feature of our corrections is that they can be used in overidentified models where the number of moment restrictions is greater than the dimension of the parameter vector. This situation is common in economic applications such as rational expectation models. Overidentification complicates the analysis by introducing an initial stage for estimating optimal weighting matrices to combine the moment conditions, and precludes the use of the existing methods. For example, Hahn and Newey's (2004) and Hahn and Kuersteiner's (2011) general bias reduction methods for nonlinear panel data models do not cover optimal two-step GMM estimators. A second distinctive feature is that our results are specifically developed for models with multidimensional nonadditive heterogeneity, whereas the previous studies focused mostly on models with additive heterogeneity captured by an scalar individual effect. Exceptions include Arellano and Hahn (2006) and Bester and Hansen (2008), who also considered multidimensional heterogeneity, but they focused on parametric likelihood-based panel models with exogenous regressors. Bai (2009) analyzed related linear panel models with exogenous regressors and multidimensional interactive individual effects. Bai's nonadditive heterogeneity allows for interaction between individual effects and unobserved factors, whereas the nonadditive heterogeneity that we consider allows for interaction between individual effects and observed regressors. The third distinctive feature of our analysis is the focus on the moments of the distribution of the individual effects or random coefficients as one of the main quantities of interest.

We illustrate the applicability of our methods with empirical and numerical examples based on the cigarette demand application of Becker, Grossman, and Murphy (1994). Here, we estimate a linear rational addictive demand model with state-specific coefficients for price and common parameters for the other regressors using a panel data set of U.S. states. We find that standard estimators that do not account for nonadditive heterogeneity by imposing a constant coefficient for price can have important biases for the common parameters, mean of the price coefficient, and demand elasticities. The analytical bias corrections are effective in removing the bias of the estimates of the mean and standard deviation of the price coefficient. Figure 1 gives a preview of the empirical results. It plots a normal approximation to the distribution of the price effect based on uncorrected and bias-corrected estimates of the mean and standard deviation

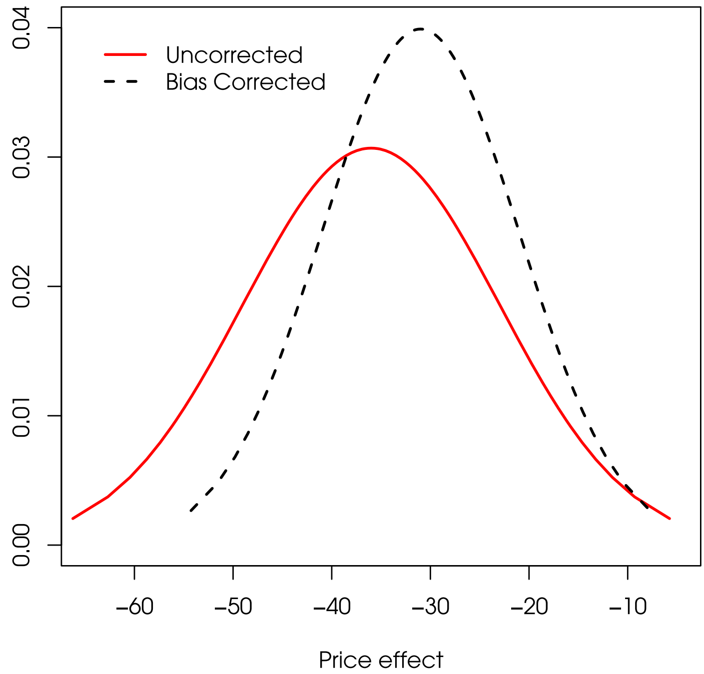

Figure 1. Normal approximation to the distribution of price effects using uncorrected (solid line) and bias-corrected (dashed line) estimates of the mean and standard deviation of the distribution of price effects. Uncorrected estimates of the mean and standard deviation are -36 and 13; bias-corrected estimates are -31 and 10.

of the distribution of the price coefficient. The figure shows that there is important heterogeneity in the price effect across states. The bias correction reduces by more than 15% the absolute value of the estimate of the mean effect and by 30% the estimate of the standard deviation.

Some of the results for the linear model are related to the recent literature on correlated random-coefficient panel models with fixed T. Graham and Powell (2012) gave identification and estimation results for average effects. Arellano and Bonhomme (2012) studied identification of the distributional characteristics of the random coefficients in exogenous linear models. None of these papers considered the case where some of the regressors have time varying endogeneity that is not captured by the random coefficients or the model is nonlinear. For nonlinear models, Chernozhukov et al. (2013) considered identification and estimation of average and quantile treatment effects. Their nonparametric and semiparametric bounds do not require large T, but they do not cover models with continuous regressors and time varying endogeneity.

The rest of the paper is organized as follows. Section 2 illustrates the type of models considered and discusses the nature of the bias in two examples. Section 3 introduces the general model and fixed effects GMM estimators. Section 4 derives the asymptotic properties of the estimators. The bias corrections and their asymptotic properties are given in Section 5. Section 6 describes the empirical and numerical examples. Section 7 concludes with a summary of the main results. Additional numerical examples, proofs, and other technical details are given in the Appendix, available in a supplementary file on the journal website, http://qeconomics.org/supp/75/supplement.pdf.

2. Motivating examples

In this section, we illustrate the nature of the bias problem with two simple examples. The first example is a linear correlated random-coefficient model with endogenous regressors. We show that averaging IV estimators applied separately to the time series of each individual is biased for the mean of the random coefficients because of the finite-sample bias of IV. The second example considers estimation of the variance of the individual coefficients in a simple setting without endogeneity. Here the sample variance of the estimators of the individual coefficients is biased because of the nonlinearity of the variance with respect to the individual coefficients. The discussion in this section is heuristic, leaving to Section 4 the specification of precise regularity conditions for the validity of the asymptotic expansions used.

2.1 *Correlated random-coefficient model with endogenous regressors*

Consider the panel model

$$y_{it} = \alpha_{0i} + \alpha_{1i} x_{it} + \varepsilon_{it} \quad (i = 1, \dots, n; t = 1, \dots, T), \tag{2.1}$$

where y_{it} is a response variable, x_{it} is an observable regressor, ε_{it} is an unobservable error term, and i and t usually index individual and time period, respectively.[2] This is a linear random-coefficient model where the effect of the regressor is heterogenous across individuals, but no restriction is imposed on the distribution of the random coefficient vector $\alpha_i := (\alpha_{0i}, \alpha_{1i})'$. The regressor can be correlated with the error term and a valid instrument $(1, z_{it})$ is available for $(1, x_{it})$ conditional on α_i, that is, $E[\varepsilon_{it} \mid \alpha_i] = 0$, $E[z_{it}\varepsilon_{it} \mid \alpha_i] = 0$, and $\mathrm{Cov}[z_{it}x_{it} \mid \alpha_i] \neq 0$. An important example of this model is the panel version of the treatment–effect model (Wooldridge (2002, Chapter 10.2.3), and Angrist and Hahn (2004)). Here, the objective is to evaluate the effect of a treatment (D) on an outcome variable (Y). The average causal effect for each level of treatment is defined as the difference between the potential outcome that the individual would obtain with and without the treatment, $Y_d - Y_0$. If individuals can choose the level of treatment, potential outcomes and levels of treatment are generally correlated. An instrumental variable Z can be used to identify the causal effect. If potential outcomes are represented as the sum of permanent individual components and transitory individual time-specific shocks, that is, $Y_{jit} = Y_{ji} + \varepsilon_{jit}$ for $j \in \{0, 1\}$, then we can write this model as a special case of (2.1) with $y_{it} = (1 - D_{it})Y_{0it} + D_{it}Y_{1it}$, $\alpha_{0i} = Y_{0i}$, $\alpha_{1i} = Y_{1i} - Y_{0i}$, $x_{it} = D_{it}$, $z_{it} = Z_{it}$, and $\varepsilon_{it} = (1 - D_{it})\varepsilon_{0it} + D_{it}\varepsilon_{1it}$.

Suppose that we are ultimately interested in $\alpha_1 := E[\alpha_{1i}]$, the mean of the random slope coefficient. We could neglect the heterogeneity and run fixed effects OLS and IV

[2] More generally, i denotes a group identifier and t indexes the observations within the group. Examples of groups include individuals, states, households, schools, and twins.

regressions in

$$y_{it} = \alpha_{0i} + \alpha_1 x_{it} + u_{it},$$

where $u_{it} = x_{it}(\alpha_{1i} - \alpha_1) + \varepsilon_{it}$ in terms of the model (2.1). In this case, OLS and IV estimate weighted means of the random coefficients; see, for example, Yitzhaki (1996) and Angrist and Krueger (1999) for OLS, and Angrist, Graddy, and Imbens (2000) for IV. Ordinary least squares puts more weight on individuals with higher variances of the regressor because they give more information about the slope, whereas IV weighs individuals in proportion to the variance of the first stage fitted values because these variances reflect the amount of information that the individuals convey about the part of the slope affected by the instrument. These weighted means are generally different from the mean coefficient because the weights can be correlated with the random coefficients.

To see how these implicit OLS and IV weighting schemes affect the estimand of the fixed-coefficient estimators, assume that the relationship between x_{it} and z_{it} is linear, that is, $x_{it} = \pi_{0i} + \pi_{1i} z_{it} + v_{it}$, $(\varepsilon_{it}, v_{it})$ is normal conditional on $(z_{it}, \alpha_i, \pi_i)$, z_{it} is independent of (α_i, π_i), and (α_i, π_i) is normal, for $\pi_i := (\pi_{0i}, \pi_{1i})'$. Then the probability limits of the OLS and IV estimators are[3]

$$\alpha_1^{\text{OLS}} = \alpha_1 + \{\text{Cov}[\varepsilon_{it}, v_{it}] + 2E[\pi_{1i}]\text{Var}[z_{it}]\text{Cov}[\alpha_{1i}, \pi_{1i}]\}/\text{Var}[x_{it}],$$
$$\alpha_1^{\text{IV}} = \alpha_1 + \text{Cov}[\alpha_{1i}, \pi_{1i}]/E[\pi_{1i}].$$

These expressions show that the OLS estimand differs from the average coefficient in the presence of endogeneity, that is, nonzero correlation between the individual time-specific error terms, or whenever the random coefficients are correlated; the IV estimand differs from the average coefficient only in the latter case.[4] In the treatment–effect model, correlation between the error terms arises in presence of endogeneity bias and correlation between the individual effects arises under Roy-type selection, that is, when individuals who experience a higher permanent effect of the treatment are relatively more prone to accept the offer of treatment. Wooldridge (2005) and Murtazashvile and Wooldridge (2005) gave sufficient conditions for consistency of standard OLS and IV fixed effects estimators. These conditions amount to $\text{Cov}[\varepsilon_{it}, v_{it}] = 0$ and $\text{Cov}[x_{it}, \alpha_{1i} | \alpha_{i0}] = 0$.

Our proposal is to estimate the mean coefficient from separate time series estimators for each individual. This strategy consists of running OLS or IV for each individual and then estimating the population moment of interest by the corresponding sample

[3]The limit of the IV estimator is obtained from a first stage equation that also imposes fixed coefficients, that is, $x_{it} = \pi_{0i} + \pi_1 z_{it} + w_{it}$, where $w_{it} = z_{it}(\pi_{1i} - \pi_1) + v_{it}$. When the first stage equation is different for each individual, the limit of the IV estimator is

$$\alpha_1^{\text{IV}} = \alpha_1 + 2E[\pi_{1i}]\text{Cov}[\alpha_{1i}, \pi_{1i}]/\{E[\pi_{1i}]^2 + \text{Var}[\pi_{1i}]\}.$$

See Theorems 2 and 3 in Angrist and Imbens (1995) for a related discussion.

[4]This feature of the IV estimator was pointed out in Angrist, Graddy, and Imbens (2000, p. 507).

moment of the individual estimators. For example, the mean of the random slope coefficient in the population is estimated by the sample average of the OLS or IV slopes. These sample moments converge to the population moments of interest as the number of individuals n and time periods T grow. However, since a different coefficient is estimated for each individual, the asymptotic distribution of the sample moments can have bias due to the incidental parameter problem (Neyman and Scott (1948)).

To illustrate the nature of this bias, consider the estimator of the mean coefficient α_1 constructed from individual time series IV estimators. In this case, the incidental parameter problem is caused by the finite-sample bias of IV. This can be explained using some expansions. Thus, assuming independence across t, standard higher-order asymptotics gives (e.g., Rilstone, Srivastava, and Ullah (1996)), as $T \to \infty$,

$$\sqrt{T}\big(\widehat{\alpha}_{1i}^{\text{IV}} - \alpha_{1i}\big) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_{it} + \frac{1}{\sqrt{T}} \beta_i + o_P\big(T^{-1/2}\big),$$

where $\psi_{it} = E[\tilde{z}_{it}\tilde{x}_{it} \mid \alpha_i, \pi_i]^{-1}\tilde{z}_{it}\varepsilon_{it}$ is the influence function of IV, $\beta_i = -E[\tilde{z}_{it}\tilde{x}_{it} \mid \alpha_i, \pi_i]^{-2}E[\tilde{z}_{it}^2\tilde{x}_{it}\varepsilon_{it} \mid \alpha_i, \pi_i]$ is the higher-order bias of IV (see, e.g., Nagar (1959), and Buse (1992)), and the variables with a tilde are deviations from their individual means, for example, $\tilde{z}_{it} = z_{it} - E[z_{it} \mid \alpha_i, \pi_i]$. In the previous expression, the first-order asymptotic distribution of the individual estimator is centered at the truth since $\sqrt{T}(\widehat{\alpha}_{1i}^{\text{IV}} - \alpha_{1i}) \to_d N(0, \sigma_i^2)$ as $T \to \infty$, where $\sigma_i^2 = E[\tilde{z}_{it}\tilde{x}_{it} \mid \alpha_i, \pi_i]^{-2}E[\tilde{z}_{it}^2\varepsilon_{it}^2 \mid \alpha_i, \pi_i]$.

Let $\widehat{\alpha}_1 = n^{-1}\sum_{i=1}^{n}\widehat{\alpha}_{1i}^{\text{IV}}$, the sample average of the IV estimators. The asymptotic distribution of $\widehat{\alpha}_1$ is not centered around α_1 in short panels or, more precisely, under asymptotic sequences where $T/\sqrt{n} \to 0$. To see this, consider the expansion for $\widehat{\alpha}_1$:

$$\sqrt{n}(\widehat{\alpha}_1 - \alpha_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\alpha_{1i} - \alpha_1) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\widehat{\alpha}_{1i}^{\text{IV}} - \alpha_{1i}\big).$$

The first term is the standard influence function for a sample mean of known elements. The second term comes from the estimation of the individual elements inside the sample mean. Assuming independence across i and combining the previous expansions,

$$\sqrt{n}(\widehat{\alpha}_1 - \alpha_1) = \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\alpha_{1i} - \alpha_1)}_{=O_P(1)} + \underbrace{\frac{1}{\sqrt{T}} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi_{it}}_{=O_P(1/\sqrt{T})} + \underbrace{\frac{\sqrt{n}}{T} \frac{1}{n} \sum_{i=1}^{n} \beta_i}_{=O(\sqrt{n}/T)} + o_P(1).$$

This expression shows that the bias term dominates the asymptotic distribution of $\widehat{\alpha}_1$ in short panels under sequences where $T/\sqrt{n} \to 0$. Averaging reduces the order of the variance of $\widehat{\alpha}_{1i}^{\text{IV}}$ without affecting the order of its bias. In this case, the estimation of the random coefficients has no first-order effect in the asymptotic variance of $\widehat{\alpha}_1$ because the second term of the expansion is of smaller order than the first term.

A potential drawback of the individual by individual time series estimation is that it might more be sensitive to weak identification problems than fixed-coefficient pooled estimation.[5] In the random-coefficient model, for example, we require that $E[\tilde{z}_{it}\tilde{x}_{it} \mid \alpha_i, \pi_i] = \pi_{1i} \neq 0$ with probability 1, that is, for all the individuals, whereas fixed-coefficient IV only requires that this condition holds on average, that is, $E[\pi_{1i}] \neq 0$. The individual estimators are, therefore, more sensitive than traditional pooled estimators to weak instruments problems. On the other hand, individual by individual estimation relaxes the exogeneity condition by conditioning on additive and nonadditive time invariant heterogeneity, that is, $E[\tilde{z}_{it}\varepsilon_{it} \mid \alpha_i, \pi_i] = 0$. Traditional fixed effects estimators only condition on additive time invariant heterogeneity. A more rigorous treatment of these identification issues is beyond the scope of this paper.

2.2 *Variance of individual coefficients*

Consider the panel model

$$y_{it} = \alpha_i + \varepsilon_{it}, \quad \varepsilon_{it} \mid \alpha_i \sim \left(0, \sigma_\varepsilon^2\right), \alpha_i \sim \left(\alpha, \sigma_\alpha^2\right) \; (t = 1, \dots, T; i = 1, \dots, n),$$

where y_{it} is an outcome variable of interest, which can be decomposed into an individual effect α_i with mean α and variance σ_α^2, and an error term ε_{it} with zero mean and variance σ_ε^2 conditional on α_i. The parameter of interest is $\sigma_\alpha^2 = \mathrm{Var}[\alpha_i]$ and its fixed effects estimator is

$$\widehat{\sigma}_\alpha^2 = (n-1)^{-1}\sum_{i=1}^n (\widehat{\alpha}_i - \widehat{\alpha})^2,$$

where $\widehat{\alpha}_i = T^{-1}\sum_{t=1}^T y_{it}$ and $\widehat{\alpha} = n^{-1}\sum_{i=1}^n \widehat{\alpha}_i$.

Let $\varphi_{\alpha_i} = (\alpha_i - \alpha)^2 - \sigma_\alpha^2$ and $\varphi_{\varepsilon_{it}} = \varepsilon_{it}^2 - \sigma_\varepsilon^2$. Assuming independence across i and t, a standard asymptotic expansion gives, as $n, T \to \infty$,

$$\sqrt{n}\left(\widehat{\sigma}_\alpha^2 - \sigma_\alpha^2\right) = \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi_{\alpha_i}}_{=O_P(1)} + \underbrace{\frac{1}{\sqrt{T}}\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T \varphi_{\varepsilon_{it}}}_{=O_P(1/\sqrt{T})} + \underbrace{\frac{\sqrt{n}}{T}\sigma_\varepsilon^2}_{=O(\sqrt{n}/T)} + o_P(1).$$

The first term corresponds to the influence function of the sample variance if the α_i's were known. The second term comes from the estimation of the α_i's. The third term is a bias term that comes from the nonlinearity of the variance in $\widehat{\alpha}_i$. The bias term dominates the expansion in short panels under sequences where $T/\sqrt{n} \to 0$. As in the previous example, the estimation of the α_i's has no first-order affect in the asymptotic variance since the second term of the expansion is of smaller order than the first term.

3. The model and estimators

We consider a general model with a finite number of moment conditions d_g. To describe it, let the data be denoted by z_{it} $(i = 1, \dots, n; t = 1, \dots, T)$. Also, let θ be a d_θ vector of common parameters, let $\{\alpha_i : 1 \leq i \leq n\}$ be a sequence of d_α vectors with the real-

[5] We thank a referee for pointing out this issue.

izations of the individual effects, and let $g(z; \theta, \alpha_i)$ be a d_g vector of functions, where $d_g \geq d_\theta + d_\alpha$.[6] The model has true parameters θ_0 and $\{\alpha_{i0} : 1 \leq i \leq n\}$ that satisfy the moment conditions

$$E[g(z_{it}; \theta_0, \alpha_{i0})] = 0 \quad (t = 1, \ldots, T; i = 1, \ldots, n),$$

where $E[\cdot]$ denotes conditional expectation with respect to the distribution of z_{it} conditional on the individual effects.

Let $\bar{E}[\cdot]$ denote the expectation taken with respect to the distribution of the individual effects. In the previous model, the ultimate quantities of interest are smooth functions of parameters and observations, which in some cases could be the parameters themselves,

$$\zeta = \bar{E}E[\zeta_i(z_{it}; \theta_0, \alpha_{i0})],$$

if $\bar{E}E|\zeta_i(z_{it}; \theta_0, \alpha_{i0})| < \infty$, or moments or other smooth functions of the individual effects,

$$\mu = \bar{E}[\mu(\alpha_{i0})],$$

if $\bar{E}|\mu(\alpha_{i0})| < \infty$. In the correlated random-coefficient example, $g(z_{it}; \theta_0, \alpha_{i0}) = z_{it}(y_{it} - \alpha_{0i0} - \alpha_{1i0}x_{it})$, $\theta = \emptyset$, $d_\theta = 0$, $d_\alpha = 2$, and $\mu(\alpha_{i0}) = \alpha_{1i0}$. In the variance of the random-coefficients example, $g(z_{it}; \theta_0, \alpha_{i0}) = (y_{it} - \alpha_{0i0})$, $\theta = \emptyset$, $d_\theta = 0$, $d_\alpha = 1$, and $\mu(\alpha_{i0}) = (\alpha_{1i0} - \bar{E}[\alpha_{1i0}])^2$.

Some more notation, which will be extensively used in the definition of the estimators and in the analysis of their asymptotic properties, is

$$\begin{aligned}
\Omega_{ji}(\theta, \alpha_i) &:= E[g(z_{it}; \theta, \alpha_i)g(z_{i,t-j}; \theta, \alpha_i)'], \quad j \in \{0, 1, 2, \ldots\}, \\
G_{\theta_i}(\theta, \alpha_i) &:= E[G_\theta(z_{it}; \theta, \alpha_i)] = E[\partial g(z_{it}; \theta, \alpha_i)/\partial\theta'], \\
G_{\alpha_i}(\theta, \alpha_i) &:= E[G_\alpha(z_{it}; \theta, \alpha_i)] = E[\partial g(z_{it}; \theta, \alpha_i)/\partial\alpha_i'],
\end{aligned}$$

where the prime denotes transpose and higher-order derivatives are denoted by adding subscripts. Here Ω_{ji} is the covariance matrix between the moment conditions for individual i at times t and $t - j$, and G_{θ_i} and G_{α_i} are time series average derivatives of the moment conditions. Analogously, for sample moments,

$$\begin{aligned}
\widehat{\Omega}_{ji}(\theta, \alpha_i) &:= T^{-1} \sum_{t=j+1}^{T} g(z_{it}; \theta, \alpha_i)g(z_{i,t-j}; \theta, \alpha_i)', \quad j \in \{0, 1, \ldots, T-1\}, \\
\widehat{G}_{\theta_i}(\theta, \alpha_i) &:= T^{-1}\sum_{t=1}^{T} G_\theta(z_{it}; \theta, \alpha_i) = T^{-1}\sum_{t=1}^{T} \partial g(z_{it}; \theta, \alpha_i)/\partial\theta', \\
\widehat{G}_{\alpha_i}(\theta, \alpha_i) &:= T^{-1}\sum_{t=1}^{T} G_\alpha(z_{it}; \theta, \alpha_i) = T^{-1}\sum_{t=1}^{T} \partial g(z_{it}; \theta, \alpha_i)/\partial\alpha_i'.
\end{aligned}$$

[6]We impose that some of the parameters are common for all the individuals to help preserve degrees of freedom in estimation of short panels with many regressors. An order condition for this model is that the number of individual-specific parameters d_α has to be less than the time dimension T.

In the sequel, the arguments of the expressions will be omitted when the functions are evaluated at the true parameter values $(\theta_0', \alpha_{i0}')'$, for example, $g(z_{it})$ means $g(z_{it}; \theta_0, \alpha_{i0})$.

In cross section and time series models, parameters defined from moment conditions are usually estimated using the two-step GMM estimator of Hansen (1982). To describe how to adapt this method to panel models with fixed effects (FE), let $\widehat{g}_i(\theta, \alpha_i) := T^{-1} \sum_{t=1}^T g(z_{it}; \theta, \alpha_i)$ and let $(\tilde{\theta}', \{\tilde{\alpha}_i'\}_{i=1}^n)'$ be some preliminary one-step FE-GMM estimator given by $(\tilde{\theta}', \{\tilde{\alpha}_i'\}_{i=1}^n)' = \arg\inf_{\{(\theta', \alpha_i')' \in \Upsilon\}_{i=1}^n} \sum_{i=1}^n \widehat{g}_i(\theta, \alpha_i)' \ \widehat{W}_i^{-1} \ \widehat{g}_i(\theta, \alpha_i)$, where $\Upsilon \subset \mathbb{R}^{d_\theta + d_\alpha}$ denotes the parameter space and $\{\widehat{W}_i : 1 \leq i \leq n\}$ is a sequence of positive definite symmetric $d_g \times d_g$ weighting matrices. The two-step FE-GMM estimator is the solution to the program

$$(\widehat{\theta}', \{\widehat{\alpha}_i'\}_{i=1}^n)' = \arg \inf_{\{(\theta', \alpha_i')' \in \Upsilon\}_{i=1}^n} \sum_{i=1}^n \widehat{g}_i(\theta, \alpha_i)' \widehat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)^{-1} \widehat{g}_i(\theta, \alpha_i),$$

where $\widehat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)$ is an estimator of the optimal weighting matrix for individual i:

$$\Omega_i = \Omega_{0i} + \sum_{j=1}^{\infty} (\Omega_{ji} + \Omega_{ji}').$$

To facilitate the asymptotic analysis, in the estimation of the optimal weighting matrix, we assume that $g(z_{it}; \theta_0, \alpha_{i0})$ is a martingale difference sequence with respect to the sigma algebra $\sigma(\alpha_i, z_{i,t-1}, z_{i,t-2}, \ldots)$, so that $\Omega_i = \Omega_{0i}$ and $\widehat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i) = \widehat{\Omega}_{0i}(\tilde{\theta}, \tilde{\alpha}_i)$. This assumption holds in rational expectation models. We do not impose this assumption to derive the limiting distribution of the one-step FE-GMM estimator.

For the subsequent analysis of the asymptotic properties of the estimator, it is convenient to consider the concentrated or profile problem. This problem is a two-step procedure. In the first step, the program is solved for the individual effects, given the value of the common parameter θ. The first-order conditions (FOC) for this stage, reparametrized conveniently as in Newey and Smith (2004), are

$$\widehat{t}_i(\theta, \widehat{\gamma}_i(\theta)) = -\begin{pmatrix} \widehat{G}_{\alpha_i}(\theta, \widehat{\alpha}_i(\theta))' \widehat{\lambda}_i(\theta) \\ \widehat{g}_i(\theta, \widehat{\alpha}_i(\theta)) + \widehat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i) \widehat{\lambda}_i(\theta) \end{pmatrix} = 0 \quad (i = 1, \ldots, n),$$

where λ_i is a d_g vector of individual Lagrange multipliers for the moment conditions and $\gamma_i := (\alpha_i', \lambda_i')'$ is an extended $(d_\alpha + d_g)$ vector of individual effects. Then the solutions to the previous equations are plugged into the original problem, leading to the first-order conditions for θ, $\widehat{s}(\widehat{\theta}) = 0$, where

$$\widehat{s}(\theta) = n^{-1} \sum_{i=1}^n \widehat{s}_i(\theta, \widehat{\gamma}_i(\theta)) = -n^{-1} \sum_{i=1}^n \widehat{G}_{\theta_i}(\theta, \widehat{\alpha}_i(\theta))' \widehat{\lambda}_i(\theta)$$

is the profile score function for θ.[7]

Fixed effects estimators of smooth functions of parameters and observations are constructed using the plug-in principle, that is, $\widehat{\zeta} = \widehat{\zeta}(\widehat{\theta})$, where

$$\widehat{\zeta}(\theta) = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} \zeta\big(z_{it}; \theta, \widehat{\alpha}_i(\theta)\big).$$

Similarly, moments of the individual effects are estimated by $\widehat{\mu} = \widehat{\mu}(\widehat{\theta})$, where

$$\widehat{\mu}(\theta) = n^{-1} \sum_{i=1}^{n} \mu\big(\widehat{\alpha}_i(\theta)\big).$$

4. Asymptotic theory for FE-GMM estimators

In this section, we analyze the properties of one-step and two-step FE-GMM estimators in large samples. We show consistency and derive the asymptotic distributions for the estimators of the individual effects, common parameters and other quantities of interest under sequences where both n and T pass to infinity with the sample size. We establish results separately for one-step and two-step estimators because the former are derived under weaker assumptions.

We make the following assumptions to show uniform consistency of the FE-GMM one-step estimator.

Condition 1 (Sampling and Asymptotics). (i) For each i, conditional on α_i, $z_i := \{z_{it} : 1 \le t \le T\}$ is a stationary mixing sequence of random vectors with strong mixing coefficients $a_i(l) = \sup_t \sup_{A \in \mathcal{A}_t^i, D \in \mathcal{D}_{t+l}^i} |P(A \cap D) - P(A)P(D)|$, where $\mathcal{A}_t^i = \sigma(\alpha_i, z_{it}, z_{i,t-1}, \ldots)$ and $\mathcal{D}_t^i = \sigma(\alpha_i, z_{it}, z_{i,t+1}, \ldots)$, such that $\sup_i |a_i(l)| \le Ca^l$ for some $0 < a < 1$ and some $C > 0$; (ii) $\{(z_i, \alpha_i) : 1 \le i \le n\}$ are independent and identically distributed across i; (iii) $n, T \to \infty$ such that $n/T \to \kappa^2$, where $0 < \kappa^2 < \infty$; and (iv) $\dim[g(\cdot; \theta, \alpha_i)] = d_g < \infty$.

For a matrix or vector A, let $|A|$ denote the Euclidean norm, that is, $|A|^2 = \text{trace}[AA']$.

Condition 2 (Regularity and Identification). (i) The vector of moment functions $g(\cdot; \theta, \alpha) = (g_1(\cdot; \theta, \alpha), \ldots, g_{d_g}(\cdot; \theta, \alpha))'$ is continuous in $(\theta, \alpha) \in \Upsilon$; (ii) the parameter space Υ is a compact, convex subset of $\mathbb{R}^{d_\theta + d_\alpha}$; (iii) $\dim(\theta, \alpha) = d_\theta + d_\alpha \le d_g$; (iv) there exists a function $M(z_{it})$ such that $|g_k(z_{it}; \theta, \alpha_i)| \le M(z_{it})$, $|\partial g_k(z_{it}; \theta, \alpha_i)/\partial(\theta, \alpha_i)| \le M(z_{it})$ for

[7]In the original parametrization, the FOC can be written as

$$n^{-1} \sum_{i=1}^{n} \widehat{G}_{\theta_i}\big(\widehat{\theta}, \widehat{\alpha}_i(\widehat{\theta})\big)' \widehat{\Omega}_i(\tilde{\theta}, \tilde{\alpha}_i)^{-} \widehat{g}_i\big(\theta, \widehat{\alpha}_i(\theta)\big) = 0,$$

where the superscript bar $(^{-})$ denotes a generalized inverse.

$k = 1, \dots, d_g$, and $\sup_i E[M(z_{it})^{4+\delta}] < \infty$ for some $\delta > 0$; and (v) there exists a deterministic sequence of symmetric finite positive definite matrices $\{W_i : 1 \le i \le n\}$ such that $\sup_{1\le i\le n} |\widehat{W}_i - W_i| \to_P 0$, and, for each $\eta > 0$,

$$\inf_i \Big[Q_i^W(\theta_0, \alpha_{i0}) - \sup_{\{(\theta,\alpha):|(\theta,\alpha)-(\theta_0,\alpha_{i0})|>\eta\}} Q_i^W(\theta, \alpha) \Big] > 0,$$

where

$$Q_i^W(\theta, \alpha_i) := -g_i(\theta, \alpha_i)' W_i^{-1} g_i(\theta, \alpha_i), \quad g_i(\theta, \alpha_i) := E\big[\widehat{g}_i(\theta, \alpha_i)\big].$$

Conditions 1(i) and (ii) impose cross sectional independence, but allow for weak time series dependence as in Hahn and Kuersteiner (2011). Conditions 1(iii) and (iv) describe the asymptotic sequences that we consider, where T and n grow at the same rate with the sample size, whereas the number of moments d_g is fixed. Condition 2 adapts standard assumptions of the GMM literature to guarantee the identification of the parameters based on time series variation for all the individuals; see Newey and McFadden (1994). The dominance and moment conditions in Condition 2(iv) are used to establish uniform consistency of the estimators of the individual effects.

Theorem 1 (Uniform Consistency of One-Step Estimators). *Suppose that Conditions 1 and 2 hold. Then, for any* $\eta > 0$,

$$\Pr\big(|\tilde{\theta} - \theta_0| \ge \eta\big) = o\big(T^{-1}\big),$$

where $\tilde{\theta} = \arg\max_{\{(\theta,\alpha_i)\in \Upsilon\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^n \widehat{Q}_i^W(\theta, \alpha_i)$ *and* $\widehat{Q}_i^W(\theta, \alpha_i) := -\widehat{g}_i(\theta, \alpha_i)' \widehat{W}_i^{-1} \widehat{g}_i(\theta, \alpha_i)$. *Also, for any* $\eta > 0$,

$$\Pr\Big(\sup_{1\le i\le n} |\tilde{\alpha}_i - \alpha_{i0}| \ge \eta\Big) = o\big(T^{-1}\big) \quad \textit{and} \quad \Pr\Big(\sup_{1\le i\le n} |\tilde{\lambda}_i| \ge \eta\Big) = o\big(T^{-1}\big),$$

where $\tilde{\alpha}_i = \arg\max_\alpha \widehat{Q}_i^W(\tilde{\theta}, \alpha)$ *and* $\tilde{\lambda}_i = -\widehat{W}_i^{-1}\widehat{g}_i(\tilde{\theta}, \tilde{\alpha}_i)$.

Let $\Sigma_{\alpha_i}^W := (G_{\alpha_i}' W_i^{-1} G_{\alpha_i})^{-1}$, $H_{\alpha_i}^W := \Sigma_{\alpha_i}^W G_{\alpha_i}' W_i^{-1}$, $P_{\alpha_i}^W := W_i^{-1} - W_i^{-1} G_{\alpha_i} H_{\alpha_i}^W$, $J_{si}^W := G_{\theta_i}' P_{\alpha_i}^W G_{\theta_i}$, and $J_s^W := \bar{E}[J_{si}^W]$. We use the following additional assumptions to derive the limiting distribution of the one-step estimator.

Condition 3 (Regularity). (i) For each i, $(\theta_0, \alpha_{i0}) \in \operatorname{int}[\Upsilon]$, and (ii) J_s^W is finite positive definite, and $\{G_{\alpha_i}' W_i^{-1} G_{\alpha_i} : 1 \le i \le n\}$ is a sequence of finite positive definite matrices, where $\{W_i : 1 \le i \le n\}$ is the sequence of matrices of Condition 2(v).

Condition 4 (Smoothness). (i) There exists a function $M(z_{it})$ such that, for $k = 1, \dots, d_g$,

$$\big|\partial^{d_1+d_2} g_k(z_{it}; \theta, \alpha_i) / \partial\theta^{d_1}\, \partial\alpha_i^{d_2}\big| \le M(z_{it}), \quad 0 \le d_1 + d_2 \le 1, \dots, 5,$$

and $\sup_i E[M(z_{it})^{5(d_\theta+d_\alpha+6)/(1-10v)+\delta}] < \infty$ for some $\delta > 0$ and $0 < v < 1/10$, and (ii) there exists $\xi_i(z_{it})$ such that $\widehat{W}_i = W_i + \sum_{t=1}^T \xi_i(z_{it})/T + R_i^W/T$, where $\max_i |R_i^W| = o_P(T^{1/2})$, $E[\xi_i(z_{it})] = 0$, and $\sup_i E[|\xi_i(z_{it})|^{20/(1-10v)+\delta}] < \infty$ for some $\delta > 0$ and $0 < v < 1/10$.

Condition 3 is the panel data analog to the standard asymptotic normality condition for GMM with cross sectional data; see Newey and McFadden (1994). Condition 4 is similar to Condition 4 in Hahn and Kuersteiner (2011), and it guarantees the existence of higher-order expansions for the GMM estimators and the uniform convergence of their remainder terms.

Let $G_{\alpha\alpha_i} := (G'_{\alpha\alpha_{i,1}}, \dots, G'_{\alpha\alpha_{i,q}})'$, where $G_{\alpha\alpha_{i,j}} = E[\partial G_{\alpha_i}(z_{it})/\partial\alpha_{i,j}]$, and let $G_{\theta\alpha_i} := (G'_{\theta\alpha_{i,1}}, \dots, G'_{\theta\alpha_{i,q}})'$, where $G_{\theta\alpha_{i,j}} = E[\partial G_{\theta_i}(z_{it})/\partial\alpha_{i,j}]$. The symbol $\otimes$ denotes the Kronecker product of matrices, I_{d_α} denotes a $d_\alpha \times d_\alpha$ identity matrix, e_j denotes a unitary d_g vector with 1 in row j, and $P^W_{\alpha_i,j}$ denotes the jth column of $P^W_{\alpha_i}$. Recall that the extended individual effect is $\gamma_i = (\alpha'_i, \lambda'_i)'$.

Lemma 1 (Asymptotic Expansion for One-Step Estimators of Individual Effects). *Under Conditions 1–4,*

$$\sqrt{T}(\tilde{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}^W_i + T^{-1/2}Q^W_{1i} + T^{-1}R^W_{2i}, \tag{4.1}$$

where $\tilde{\gamma}_{i0} := \tilde{\gamma}_i(\theta_0)$,

$$\tilde{\psi}^W_i = -\begin{pmatrix} H^W_{\alpha_i} \\ P^W_{\alpha_i} \end{pmatrix} T^{-1/2}\sum_{t=1}^{T} g(z_{it}) \xrightarrow{d} N\big(0, V^W_i\big),$$

$n^{-1/2}\sum_{i=1}^n \tilde{\psi}^W_i \xrightarrow{d} N(0, \bar{E}[V^W_i])$, $n^{-1}\sum_{i=1}^n Q^W_{1i} \xrightarrow{p} \bar{E}[B^W_{\gamma_i}]$, $B^W_{\gamma_i} = B^{W,I}_{\gamma_i} + B^{W,G}_{\gamma_i} + B^{W,1S}_{\gamma_i}$, *and* $\sup_{1\le i\le n} R^W_{2i} = o_P(\sqrt{T})$ *for*

$$V^W_i = \begin{pmatrix} H^W_{\alpha_i} \\ P^W_{\alpha_i} \end{pmatrix} \Omega_i\big(H^{W'}_{\alpha_i}, P^W_{\alpha_i}\big),$$

$$B^{W,I}_{\gamma_i} = \begin{pmatrix} B^{W,I}_{\alpha_i} \\ B^{W,I}_{\lambda_i} \end{pmatrix}$$

$$= \begin{pmatrix} H^W_{\alpha_i} \\ P^W_{\alpha_i} \end{pmatrix} \left(\sum_{j=-\infty}^{\infty} E\big[G_{\alpha_i}(z_{it})H^W_{\alpha_i} g(z_{i,t-j})\big] - \sum_{j=1}^{d_\alpha} G_{\alpha\alpha_{i,j}} H^W_{\alpha_i} \Omega_i H^{W'}_{\alpha_i}/2 \right),$$

$$B^{W,G}_{\gamma_i} = \begin{pmatrix} B^{W,G}_{\alpha_i} \\ B^{W,G}_{\lambda_i} \end{pmatrix} = \begin{pmatrix} -\Sigma^W_{\alpha_i} \\ H^{W'}_{\alpha_i} \end{pmatrix} \sum_{j=-\infty}^{\infty} E\big[G_{\alpha_i}(z_{it})' P^W_{\alpha_i} g(z_{i,t-j})\big],$$

$$B^{W,1S}_{\gamma_i} = \begin{pmatrix} B^{W,1S}_{\alpha_i} \\ B^{W,1S}_{\lambda_i} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma^W_{\alpha_i} \\ -H^{W'}_{\alpha_i} \end{pmatrix} \left(\sum_{j=1}^{d_\alpha} G'_{\alpha\alpha_{i,j}} P^W_{\alpha_i} \Omega_i H^{W'}_{\alpha_i}/2 + \sum_{j=1}^{d_g} G'_{\alpha\alpha_i}(I_{d_\alpha} \otimes e_j) H^W_{\alpha_i} \Omega_i P^W_{\alpha_i,j}/2 \right)$$

$$+ \begin{pmatrix} H^W_{\alpha_i} \\ P^W_{\alpha_i} \end{pmatrix} \sum_{j=-\infty}^{\infty} E\big[\xi_i(z_{it}) P^W_{\alpha_i} g(z_{i,t-j})\big].$$

Theorem 2 (Limit Distribution of One-Step Estimators of Common Parameters). *Under Conditions 1–4,*

$$\sqrt{nT}(\tilde{\theta}-\theta_0) \xrightarrow{d} -\left(J_s^W\right)^{-1} N\left(\kappa B_s^W, V_s^W\right),$$

where

$$J_s^W = \bar{E}\left[G_{\theta_i}' P_{\alpha_i}^W G_{\theta_i}\right],$$
$$V_s^W = \bar{E}\left[G_{\theta_i}' P_{\alpha_i}^W \Omega_i P_{\alpha_i}^W G_{\theta_i}\right],$$
$$B_s^W = \bar{E}\left[B_{si}^{W,B} + B_{si}^{W,C} + B_{si}^{W,V}\right]$$

and

$$B_{si}^{W,B} = -G_{\theta_i}'\left(B_{\lambda_i}^{W,I} + B_{\lambda_i}^{W,G} + B_{\lambda_i}^{W,1S}\right),$$

$$B_{si}^{W,C} = \sum_{j=-\infty}^{\infty} E\left[G_{\theta_i}(z_{it})' P_{\alpha_i}^W g_i(z_{i,t-j})\right],$$

$$B_{si}^{W,V} = -\sum_{j=1}^{d_\alpha} G_{\theta\alpha_{i,j}}' P_{\alpha_i}^W \Omega_i H_{\alpha_i}^{W\prime}/2 - \sum_{j=1}^{d_g} G_{\theta\alpha_i}'(I_{d_\alpha} \otimes e_j) H_{\alpha_i}^W \Omega_i P_{\alpha_i,j}/2.$$

The expressions for $B_{\lambda_i}^{W,I}$, $B_{\lambda_i}^{W,G}$, and $B_{\lambda_i}^{W,1S}$ are given in Lemma 1.

The source of the bias is the nonzero expectation of the profile score of θ at the true parameter value produced by the substitution of the unobserved individual effects by sample estimators. These estimators converge to their true parameter value at a rate $\sqrt{T}$, which is slower than $\sqrt{nT}$, the rate of convergence of the estimator of the common parameter. Intuitively, the rate for $\tilde{\gamma}_{i0}$ is $\sqrt{T}$ because only the T observations for individual i convey information about γ_{i0}. In nonlinear and dynamic models, the slow convergence of the estimator of the individual effect introduces bias in the estimators of the rest of parameters. The expression of this bias can be explained with an expansion of the score around the true value of the individual effects[8]:

$$\begin{aligned} E\left[\widehat{s}_i^W(\theta_0, \tilde{\gamma}_{i0})\right] &= E\left[\widehat{s}_i^W\right] + E\left[\widehat{s}_{\gamma i}^W\right]' E[\tilde{\gamma}_{i0} - \gamma_{i0}] + E\left[\left(\widehat{s}_{\gamma i}^W - E\left[\widehat{s}_{\gamma i}^W\right]\right)'(\tilde{\gamma}_{i0} - \gamma_{i0})\right] \\ &\quad + E\left[\sum_{j=1}^{d_\alpha + d_g} (\tilde{\gamma}_{i0,j} - \gamma_{i0,j}) E\left[\widehat{s}_{\gamma\gamma i}^W\right](\tilde{\gamma}_{i0} - \gamma_{i0})\right] \Big/ 2 + o\left(T^{-1}\right) \\ &= 0 + B_s^{W,B}/T + B_s^{W,C}/T + B_s^{W,V}/T + o\left(T^{-1}\right). \end{aligned}$$

[8]Using the notation introduced in Section 3, the score is

$$\widehat{s}^W(\theta_0) = n^{-1} \sum_{i=1}^n \widehat{s}_i^W(\theta_0, \tilde{\gamma}_{i0}) = -n^{-1} \sum_{i=1}^n \widehat{G}_{\theta_i}(\theta_0, \tilde{\alpha}_{i0})' \tilde{\lambda}_{i0},$$

where $\tilde{\gamma}_{i0} = (\tilde{\alpha}_{i0}', \tilde{\lambda}_{i0}')$ is the solution to

$$\widehat{t}_i^W(\theta_0, \tilde{\gamma}_{i0}) = -\begin{pmatrix} \widehat{G}_{\alpha_i}(\theta_0, \tilde{\alpha}_{i0})' \tilde{\lambda}_{i0} \\ \widehat{g}_i(\theta_0, \tilde{\alpha}_{i0}) + W_i \tilde{\lambda}_{i0} \end{pmatrix} = 0.$$

This expression shows that the bias has the same three components as in the maximum likelihood estimator (MLE) case; see Hahn and Newey (2004). The first component, $B_s^{W,B}$, comes from the higher-order bias of the estimator of the individual effects. The second component, $B_s^{W,C}$, is a correlation term and is present because individual effects and common parameters are estimated using the same observations. The third component, $B_s^{W,V}$, is a variance term. The bias of the individual effects, $B_s^{W,B}$, can be further decomposed into three terms that correspond to the asymptotic bias for a GMM estimator with the optimal score ($B_\lambda^{W,I}$) when W is used as the weighting function, the bias arising from estimation of G_{α_i} ($B_\lambda^{W,G}$), and the bias arising from not using an optimal weighting matrix ($B_\lambda^{W,1S}$).

We use the following condition to show the consistency of the two-step FE-GMM estimator.

Condition 5 (Smoothness, Regularity, and Martingale). (i) There exists a function $M(z_{it})$ such that $|g_k(z_{it};\theta,\alpha_i)| \le M(z_{it})$, $|\partial g_k(z_{it};\theta,\alpha_i)/\partial(\theta,\alpha_i)| \le M(z_{it})$ for $k=1,\dots,d_g$, and $\sup_i E[M(z_{it})^{10(d_\theta+d_\alpha+6)/(1-10v)+\delta}] < \infty$ for some $\delta > 0$ and $0 < v < 1/10$; (ii) $\{\Omega_i : 1 \le i \le n\}$ is a sequence of finite positive definite matrices; and (iii) for each i, $g(z_{it};\theta_0,\alpha_{i0})$ is a martingale difference sequence with respect to $\sigma(\alpha_i, z_{i,t-1}, z_{i,t-2},\dots)$.

Conditions 5(i) and (ii) are used to establish the uniform consistency of the estimators of the individual weighting matrices. Condition 5(iii) is convenient to simplify the expressions of the optimal weighting matrices. It holds, for example, in rational expectation models that commonly arise in economic applications.

Theorem 3 (Uniform Consistency of Two-Step Estimators). *Suppose that Conditions 1, 2, 3, and 5 hold. Then, for any* $\eta > 0$,

$$\Pr\big(|\widehat{\theta}-\theta_0| \ge \eta\big) = o\big(T^{-1}\big),$$

where $\widehat{\theta} = \arg\max_{\{(\theta',\alpha_i')\}_{i=1}^n \in \Upsilon} \sum_{i=1}^n \widehat{Q}_i^\Omega(\theta,\alpha_i)$ *and* $\widehat{Q}_i^\Omega(\theta,\alpha_i) := -\widehat{g}_i(\theta,\alpha_i)'\widehat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)^{-1}\widehat{g}_i(\theta,\alpha_i)$. *Also, for any* $\eta > 0$,

$$\Pr\Big(\sup_{1\le i\le n} |\widehat{\alpha}_i - \alpha_0| \ge \eta\Big) = o\big(T^{-1}\big) \quad \textit{and} \quad \Pr\Big(\sup_{1\le i\le n} |\widehat{\lambda}_i| \ge \eta\Big) = o\big(T^{-1}\big),$$

where $\widehat{\alpha}_i = \arg\max_\alpha \widehat{Q}_i^\Omega(\widehat{\theta},\alpha)$ *and* $\widehat{g}_i(\widehat{\theta},\widehat{\alpha}_i) + \widehat{\Omega}_i(\tilde{\theta},\tilde{\alpha}_i)\widehat{\lambda}_i = 0$.

We replace Condition 4 by the following condition to obtain the limit distribution of the two-step estimator.

Condition 6 (Smoothness). There exists some $M(z_{it})$ such that, for $k=1,\dots,d_g$,

$$\big|\partial^{d_1+d_2} g_k(z_{it};\theta,\alpha_i)/\partial\theta^{d_1}\,\partial\alpha_i^{d_2}\big| \le M(z_{it}), \quad 0 \le d_1+d_2 \le 1,\dots,5,$$

and $\sup_i E[M(z_{it})^{10(d_\theta+d_\alpha+6)/(1-10v)+\delta}] < \infty$ for some $\delta > 0$ and $0 < v < 1/10$.

Condition 6 guarantees the existence of higher-order expansions for the estimators of the weighting matrices and uniform convergence of their remainder terms. Conditions 5 and 6 are stronger versions of Conditions 2(iv), 2(v), and 4. They are presented separately because they are only needed when there is a first stage where the weighting matrices are estimated.

Let $\Sigma_{\alpha_i} := (G'_{\alpha_i}\Omega_i^{-1}G_{\alpha_i})^{-1}$, $H_{\alpha_i} := \Sigma_{\alpha_i}G'_{\alpha_i}\Omega_i^{-1}$, and $P_{\alpha_i} := \Omega_i^{-1} - \Omega_i^{-1}G_{\alpha_i}H_{\alpha_i}$.

Lemma 2 (Asymptotic Expansion for Two-Step Estimators of Individual Effects). *Under Conditions* 1–5,

$$\sqrt{T}(\widehat{\gamma}_{i0} - \gamma_{i0}) = \tilde{\psi}_i + T^{-1/2}B_{\gamma_i} + T^{-1}R_{2i}, \tag{4.2}$$

where $\widehat{\gamma}_{i0} := \widehat{\gamma}_i(\theta_0)$,

$$\tilde{\psi}_i = -\begin{pmatrix} H_{\alpha_i} \\ P_{\alpha_i} \end{pmatrix} T^{-1/2}\sum_{t=1}^{T} g(z_{it}) \xrightarrow{d} N(0, V_i),$$

$n^{-1/2}\sum_{i=1}^{n}\tilde{\psi}_i \xrightarrow{d} N(0, \bar{E}[V_i])$, $B_{\gamma_i} = B_{\gamma_i}^I + B_{\gamma_i}^G + B_{\gamma_i}^\Omega + B_{\gamma_i}^W$, *and* $\sup_{1\le i\le n} R_{2i} = o_P(\sqrt{T})$, *with, for* $\Omega_{\alpha_{i,j}} = \partial\Omega_{\alpha_i}/\partial\alpha_{i,j}$,

$$V_i = \operatorname{diag}(\Sigma_{\alpha_i}, P_{\alpha_i}),$$

$$B_{\gamma_i}^I = \begin{pmatrix} B_{\alpha_i}^I \\ B_{\lambda_i}^I \end{pmatrix} = \begin{pmatrix} H_{\alpha_i} \\ P_{\alpha_i} \end{pmatrix}\left(-\sum_{j=1}^{d_\alpha} G_{\alpha\alpha_{i,j}}\Sigma_{\alpha_i}/2 + E\big[G_{\alpha_i}(z_{it})H_{\alpha_i}g(z_{i,t-j})\big]\right),$$

$$B_{\gamma_i}^G = \begin{pmatrix} B_{\alpha_i}^G \\ B_{\lambda_i}^G \end{pmatrix} = \begin{pmatrix} -\Sigma_{\alpha_i} \\ H'_{\alpha_i} \end{pmatrix}\sum_{j=0}^{\infty} E\big[G_{\alpha_i}(z_{it})'P_{\alpha_i}g(z_{i,t-j})\big],$$

$$B_{\gamma_i}^\Omega = \begin{pmatrix} B_{\alpha_i}^\Omega \\ B_{\lambda_i}^\Omega \end{pmatrix} = \begin{pmatrix} H_{\alpha_i} \\ P_{\alpha_i} \end{pmatrix}\sum_{j=0}^{\infty} E\big[g(z_{it})g(z_{it})'P_{\alpha_i}g(z_{i,t-j})\big],$$

$$B_{\gamma_i}^W = \begin{pmatrix} B_{\alpha_i}^W \\ B_{\lambda_i}^W \end{pmatrix} = \begin{pmatrix} H_{\alpha_i} \\ P_{\alpha_i} \end{pmatrix}\sum_{j=1}^{d_\alpha} \Omega_{\alpha_{i,j}}\big(H_{\alpha_{i,j}}^{W'} - H'_{\alpha_{i,j}}\big).$$

Theorem 4 (Limit Distribution for Two-Step Estimators of Common Parameters). *Under Conditions* 1–6,

$$\sqrt{nT}(\widehat{\theta} - \theta_0) \xrightarrow{d} -J_s^{-1}N(\kappa B_s, J_s),$$

where $J_s = \bar{E}[G'_{\theta_i}P_{\alpha_i}G_{\theta_i}]$, $B_s = \bar{E}[B_{si}^B + B_{si}^C]$, $B_{si}^B = -G'_{\theta_i}[B_{\lambda_i}^I + B_{\lambda_i}^G + B_{\lambda_i}^\Omega + B_{\lambda_i}^W]$, *and* $B_{si}^C = \sum_{j=0}^{\infty}E[G_{\theta_i}(z_{it})'P_{\alpha_i}g(z_{i,t-j})]$. *The expressions for* $B_{\lambda_i}^I$, $B_{\lambda_i}^G$, $B_{\lambda_i}^\Omega$, *and* $B_{\lambda_i}^W$ *are given in Lemma* 2.

Theorem 4 establishes that one iteration of the GMM procedure not only improves asymptotic efficiency by reducing the variance of the influence function, but also removes the variance and nonoptimal weighting matrices components from the bias. The higher-order bias of the estimator of the individual effects, B_λ^B, now has four components, as in Newey and Smith (2004). These components correspond to the asymptotic bias for a GMM estimator with the optimal score (B_λ^I), the bias arising from estimation of G_{α_i} (B_λ^G), the bias arising from estimation of Ω_i (B_λ^Ω), and the bias arising from the choice of the preliminary first step estimator (B_λ^W). An additional iteration of the GMM estimator removes the term B_λ^W.

The general procedure for deriving the asymptotic distribution of the FE-GMM estimators consists of several expansions. First, we derive higher-order asymptotic expansions for the estimators of the individual effects, with the common parameter fixed at its true value θ_0. Next, we obtain the asymptotic distribution for the profile score of the common parameter at θ_0 using the expansions of the estimators of the individual effects. Finally, we derive the asymptotic distribution of the estimator for the common parameter by multiplying the asymptotic distribution of the score by the limit profile Jacobian matrix. This procedure is detailed in the Appendix. Here we characterize the asymptotic bias in a linear correlated random-coefficient model with endogenous regressors. Motivated by the numerical and empirical examples that follow, we consider a model where only the variables with a common parameter are endogenous and we allow for the moment conditions not to be martingale difference sequences.

Example (Correlated Random-Coefficient Model With Endogenous Regressors). We consider a simplified version of the models in the empirical and numerical examples. The notation is the same as in the theorems discussed above. The moment condition is

$$g(z_{it}; \theta, \alpha_i) = w_{it}\big(y_{it} - x_{1it}'\alpha_i - x_{2it}'\theta\big),$$

where $w_{it} = (x_{1it}', w_{2it}')'$ and $z_{it} = (x_{1it}', x_{2it}', w_{2it}', y_{it})'$. That is, only the regressors with common coefficients are endogenous. Let $\varepsilon_{it} = y_{it} - x_{1it}'\alpha_{i0} - x_{2it}'\theta_0$. To simplify the expressions for the bias, we assume that $\varepsilon_{it} \mid w_i, \alpha_i \sim$ i.i.d. $(0, \sigma_\varepsilon^2)$ and $E[x_{2it}\varepsilon_{i,t-j} \mid w_i, \alpha_i] = E[x_{2it}\varepsilon_{i,t-j}]$ for $w_i = (w_{i1}, \ldots, w_{iT})'$ and $j \in \{0, \pm 1, \ldots\}$. Under these conditions, the optimal weighted matrices are proportional to $E[w_{it}w_{it}']$, which do not depend on θ_0 and α_{i0}. We can, therefore, obtain the optimal GMM estimator in one step using the sample averages $T^{-1}\sum_{t=1}^T w_{it}w_{it}'$ to estimate the optimal weighting matrices.

In this model, it is straightforward to see that the estimators of the individual effects have no bias, that is, $B_{\gamma_i}^{W,I} = B_{\gamma_i}^{W,G} = B_{\gamma_i}^{W,1S} = 0$. By linearity of the first-order conditions in θ and α_i, $B_{si}^{W,V} = 0$. The only source of bias is the correlation between the estimators of θ and α_i. After some straightforward algebra, this bias simplifies to

$$B_{si}^{W,C} = -(d_g - d_\alpha)\sum_{j=-\infty}^{\infty} E[x_{2it}\varepsilon_{i,t-j}].$$

For the limit Jacobian, we find

$$J_s^W = \bar{E}\big\{E\big[\tilde{x}_{2it}\tilde{w}_{2it}'\big]E\big[\tilde{w}_{2it}\tilde{w}_{2it}'\big]^{-1}E\big[\tilde{w}_{2it}\tilde{x}_{2it}'\big]\big\},$$

where variables with a tilde indicate residuals of population linear projections of the corresponding variable on x_{1it}, for example, $\tilde{x}_{2it} = x_{2it} - E[x_{2it}x_{1it}']E[x_{1it}x_{1it}']^{-1}x_{1it}$. The expression of the bias is

$$\mathcal{B}(\theta_0) = -(d_g - d_\alpha)\big(J_s^W\big)^{-1}\bar{E}\sum_{j=-\infty}^{\infty} E\big[\tilde{x}_{2it}\big(\tilde{y}_{i,t-j} - \tilde{x}'_{2i,t-j}\theta_0\big)\big]. \tag{4.3}$$

□

In random-coefficient models the ultimate quantities of interest are often functions of the data, model parameters, and individual effects. The following corollaries characterize the asymptotic distributions of the fixed effects estimators of these quantities. The first corollary applies to averages of functions of the data and individual effects such as average partial effects and average derivatives in nonlinear models, and average elasticities in linear models with variables in levels. Section 6 gives an example of these elasticities. The second corollary applies to averages of smooth functions of the individual effects including means, variances, and other moments of the distribution of these effects. Sections 2 and 6 give examples of these functions. We state the results only for estimators constructed from two-step estimators of the common parameters and individual effects. Similar results apply to estimators constructed from one-step estimators. Both corollaries follow from Lemma 2 and Theorem 4 by the delta method.

Corollary 1 (Asymptotic Distribution for Fixed Effects Averages). *Let $\zeta(z;\theta,\alpha_i)$ be a twice continuously differentiable function in its second and third argument such that* $\inf_i \operatorname{Var}[\zeta(z_{it})] > 0$, $\bar{E}E[\zeta(z_{it})^2] < \infty$, $\bar{E}E|\zeta_\alpha(z_{it})|^2 < \infty$, *and* $\bar{E}E|\zeta_\theta(z_{it})|^2 < \infty$, *where the subscripts on ζ denote partial derivatives. Then, under the conditions of Theorem 4, for some deterministic sequence $r_{nT} \to \infty$ such that $r_{nT} = O(\sqrt{nT})$,*

$$r_{nT}(\widehat{\zeta} - \zeta - B_\zeta/T) \xrightarrow{d} N(0, V_\zeta),$$

where $\zeta = \bar{E}E[\zeta(z_{it})]$,

$$B_\zeta = \bar{E}E\Bigg[-\sum_{j=0}^{\infty}\zeta_{\alpha_i}(z_{it})'H_{\alpha_i}g(z_{i,t-j}) + \zeta_{\alpha_i}(z_{it})'B_{\alpha_i} + \sum_{j=1}^{d_\alpha}\zeta_{\alpha\alpha_{i,j}}(z_{it})'\Sigma_{\alpha_i}/2 - \zeta_\beta(z_{it})'J_s^{-1}B_s\Bigg]$$

for $B_{\alpha_i} = B_{\alpha_i}^I + B_{\alpha_i}^G + B_{\alpha_i}^\Omega + B_{\alpha_i}^W$, *and for* $r^2 = \lim_{n,T\to\infty} r_{nT}^2/(nT)$,

$$V_\zeta = r^2\bar{E}E\big[\zeta_{\alpha_i}(z_{it})'\Sigma_{\alpha_i}\zeta_{\alpha_i}(z_{it}) + \zeta_\theta(z_{it})'J_s^{-1}\zeta_\theta(z_{it})\big] + \lim_{n,T\to\infty}\frac{r_{nT}^2}{n}\bar{E}E\Bigg[\bigg(\frac{1}{T}\sum_{t=1}^{T}\big(\zeta(z_{it}) - \zeta\big)\bigg)^2\Bigg].$$

Corollary 2 (Asymptotic Distribution for Smooth Functions of Individual Effects). *Let $\mu(\alpha_i)$ be a twice differentiable function such that $\bar{E}[\mu(\alpha_{i0})^2] < \infty$ and $\bar{E}|\mu_\alpha(\alpha_{i0})|^2 < \infty$,* *where the subscripts on* μ *denote partial derivatives. Then, under the conditions of Theorem 4,*

$$\sqrt{n}(\widehat{\mu}-\mu)\xrightarrow{d}N(\kappa B_{\mu},V_{\mu}),$$

where $\mu=\bar{E}[\mu(\alpha_{i0})]$,

$$B_{\mu}=\bar{E}\left[\mu_{\alpha_i}(\alpha_{i0})'B_{\alpha_i}+\sum_{j=1}^{d_{\alpha}}\mu_{\alpha\alpha_{i,j}}(\alpha_{i0})'\Sigma_{\alpha_i}/2\right]$$

for $B_{\alpha_i}=B^I_{\alpha_i}+B^G_{\alpha_i}+B^{\Omega}_{\alpha_i}+B^W_{\alpha_i}$, *and* $V_{\mu}=\bar{E}[(\mu(\alpha_{i0})-\mu)^2]$.

The convergence rate r_{nT} in Corollary 1 depends on the function $\zeta(z;\theta,\alpha_i)$. For example, $r_{nT}=\sqrt{nT}$ for functions that do not depend on α_i such as $\zeta(z;\theta,\alpha_i)=c'\theta$, where c is a known d_θ vector. In general, $r_{nT}=\sqrt{n}$ for functions that depend on α_i. In this case $r^2=0$ and the first two terms of V_ζ drop out. Corollary 2 is an important special case of Corollary 1. We present it separately because the asymptotic bias and variance have simplified expressions.

5. Bias corrections

The FE-GMM estimators of common parameters, while consistent, have bias in the asymptotic distributions under sequences where n and T grow at the same rate. These sequences provide a good approximation to the finite-sample behavior of the estimators in empirical applications where the time dimension is moderately large. The presence of bias invalidates any asymptotic inference because the bias is of the same order as the standard deviation. In this section, we describe bias correction methods to adjust the asymptotic distribution of the FE-GMM estimators of the common parameter and smooth functions of the data, model parameters, and individual effects. All the corrections considered are analytical. Alternative corrections based on variations of the jackknife can be implemented using the approaches described in Hahn and Newey (2004) and Dhaene and Jochmans (2010).[9]

We consider three analytical methods that differ in whether the bias is corrected from the estimator or from the first-order conditions, and in whether the correction is one-step or iterated for methods that correct the bias from the estimator. All these methods reduce the order of the asymptotic bias without increasing the asymptotic variance. They are based on analytical estimators of the bias of the profile score B_s and the profile Jacobian matrix J_s. Since these quantities include cross sectional and time series means $\bar{E}$ and E evaluated at the true parameter values for the common parameter and individual effects, they are estimated by the corresponding cross sectional and time series sample averages evaluated at the FE-GMM estimates. Thus, for any function of the data,

[9] Hahn, Kuersteiner, and Newey (2004) showed that analytical, bootstrap, and jackknife bias correction methods are asymptotically equivalent up to third order for MLE. We conjecture that the same result applies to GMM estimators, but the proof is beyond the scope of this paper.

common parameter and individual effects $f_{it}(\theta, \alpha_i)$, let $\widehat{f}_{it}(\theta) = f_{it}(\theta, \widehat{\alpha}_i(\theta))$, $\widehat{f}_i(\theta) = \widehat{E}[\widehat{f}_{it}(\theta)] = T^{-1}\sum_{t=1}^{T}\widehat{f}_{it}(\theta)$, and $\widehat{f}(\theta) = \widehat{\bar{E}}[\widehat{f}_i(\theta)] = n^{-1}\sum_{i=1}^{n}\widehat{f}_i(\theta)$. Next, define $\widehat{\Sigma}_{\alpha_i}(\theta) = [\widehat{G}_{\alpha_i}(\theta)'\widehat{\Omega}_i^{-1}\widehat{G}_{\alpha_i}(\theta)]^{-1}$, $\widehat{H}_{\alpha_i}(\theta) = \widehat{\Sigma}_{\alpha_i}(\theta)\widehat{G}_{\alpha_i}(\theta)'\widehat{\Omega}_i^{-1}$, and $\widehat{P}_{\alpha_i}(\theta) = \widehat{\Omega}_i^{-1}\widehat{G}_{\alpha_i}(\theta)\widehat{H}_{\alpha_i}(\theta)$. To simplify the presentation, we only give explicit formulas for FE-GMM three-step estimators in the main text. We give the expressions for one- and two-step estimators in the Appendix. Let

$$\widehat{\mathcal{B}}(\theta) = -\widehat{J}_s(\theta)^{-1}\widehat{B}_s(\theta), \qquad \widehat{B}_s(\theta) = \widehat{\bar{E}}\big[\widehat{B}_{si}^B(\theta) + \widehat{B}_{si}^C(\theta)\big],$$
$$\widehat{J}_s(\theta) = \widehat{\bar{E}}\big[\widehat{G}_{\theta_i}(\theta)'\widehat{P}_{\alpha_i}(\theta)\widehat{G}_{\theta_i}(\theta)\big],$$

where $\widehat{B}_{si}^B(\theta) = -\widehat{G}_{\theta_i}(\theta)'[\widehat{B}_{\lambda_i}^I(\theta) + \widehat{B}_{\lambda_i}^G(\theta) + \widehat{B}_{\lambda_i}^\Omega(\theta) + \widehat{B}_{\lambda_i}^W(\theta)]$,

$$\widehat{B}_{\lambda_i}^I(\theta) = -\widehat{P}_{\alpha_i}(\theta)\sum_{j=1}^{d_\alpha}\widehat{G}_{\alpha\alpha_{i,j}}(\theta)\widehat{\Sigma}_{\alpha_i}(\theta)/2$$
$$\qquad + \widehat{P}_{\alpha_i}(\theta)\sum_{j=0}^{\ell}T^{-1}\sum_{t=j+1}^{T}\widehat{G}_{\alpha_{it}}(\theta)\widehat{H}_{\alpha_i}(\theta)\widehat{g}_{i,t-j}(\theta),$$
$$\widehat{B}_{\lambda_i}^G(\theta) = \widehat{H}_{\alpha_i}(\theta)'\sum_{j=0}^{\infty}T^{-1}\sum_{t=j+1}^{T}\widehat{G}_{\alpha_{it}}(\theta)'\widehat{P}_{\alpha_i}(\theta)\widehat{g}_{i,t-j}(\theta),$$
$$\widehat{B}_{\lambda_i}^\Omega(\theta) = \widehat{P}_{\alpha_i}(\theta)\sum_{j=0}^{\ell}T^{-1}\sum_{t=j+1}^{T}\widehat{g}_{it}(\theta)\widehat{g}_{it}(\theta)'\widehat{P}_{\alpha_i}(\theta)\widehat{g}_{i,t-j}(\theta),$$

and $\widehat{B}_{si}^C(\theta) = T^{-1}\sum_{j=0}^{\ell}\sum_{t=j+1}^{T}\widehat{G}_{\theta_{it}}(\theta)'\widehat{P}_{\alpha_i}(\theta)\widehat{g}_{i,t-j}(\theta)$. In the previous expressions, the spectral time series averages that involve an infinite number of terms are trimmed. The trimming parameter ℓ is a positive bandwidth that needs to be chosen such that $\ell \to \infty$ and $\ell/T \to 0$ as $T \to \infty$ (Hahn and Kuersteiner (2011)).

The one-step correction of the estimator subtracts an estimator of the expression of the asymptotic bias from the estimator of the common parameter. Using the expressions defined above evaluated at $\widehat{\theta}$, the bias-corrected estimator is

$$\widehat{\theta}^{\mathrm{BC}} = \widehat{\theta} - \widehat{\mathcal{B}}(\widehat{\theta})/T. \tag{5.1}$$

This bias correction is straightforward to implement because it only requires one optimization. The iterated correction is equivalent to solving the nonlinear equation

$$\widehat{\theta}^{\mathrm{IBC}} = \widehat{\theta} - \widehat{\mathcal{B}}\big(\widehat{\theta}^{\mathrm{IBC}}\big)/T. \tag{5.2}$$

When $\theta + \widehat{\mathcal{B}}(\theta)$ is invertible in θ, it is possible to obtain a closed-form solution to the previous equation.[10] Otherwise, an iterative procedure is needed. The score bias-corrected (SBC) estimator is the solution to the estimating equation

$$\widehat{s}\big(\widehat{\theta}^{\mathrm{SBC}}\big) - \widehat{B}_s\big(\widehat{\theta}^{\mathrm{SBC}}\big)/T = 0. \tag{5.3}$$

[10]See MacKinnon and Smith (1998) for a comparison of one-step and iterated bias correction methods.

This procedure, while computationally more intensive, has the attractive feature that both estimator and bias are obtained simultaneously. Hahn and Newey (2004) showed that fully iterated bias-corrected (IBC) estimators solve approximated bias-corrected first-order conditions. The IBC and SBC are equivalent if the first-order conditions are linear in θ.

Example (Correlated Random-Coefficient Model With Endogenous Regressors). The previous methods can be illustrated in the correlated random-coefficient model example in Section 4. Here, the fixed effects GMM estimators have closed forms

$$\widehat{\alpha}_i(\theta) = \left(\sum_{t=1}^{T} x_{1it}x_{1it}'\right)^{-1} \sum_{t=1}^{T} x_{1it}\bigl(y_{it} - x_{2it}'\theta\bigr)$$

and

$$\widehat{\theta} = \bigl(\widehat{J}_s^W\bigr)^{-1} \sum_{i=1}^{n} \left[\sum_{t=1}^{T} \tilde{x}_{2it}\tilde{w}_{2it}' \left(\sum_{t=1}^{T} \tilde{w}_{2it}\tilde{w}_{2it}'\right)^{-1} \sum_{t=1}^{T} \tilde{w}_{2it}\tilde{y}_{it}\right],$$

where $\widehat{J}_s^W = \sum_{i=1}^n [\sum_{t=1}^T \tilde{x}_{2it}\tilde{w}_{2it}' (\sum_{t=1}^T \tilde{w}_{2it}\tilde{w}_{2it}')^{-1} \sum_{t=1}^T \tilde{w}_{2it}\tilde{x}_{2it}']$ and the variables with a tilde now indicate residuals of sample linear projections of the corresponding variable on x_{1it}, for example, $\tilde{x}_{2it} = x_{2it} - \sum_{t=1}^T x_{2it}x_{1it}'(\sum_{t=1}^T x_{1it}x_{1it}')^{-1}x_{1it}$.

We can estimate the bias of $\widehat{\theta}$ from the analytic formula in expression (4.3), replacing population by sample moments, replacing θ_0 by $\widehat{\theta}$, and trimming the number of terms in the spectral expectation:

$$\widehat{\mathcal{B}}(\widehat{\theta}) = -(d_g - d_\alpha)\bigl(\widehat{J}_s^W\bigr)^{-1} \sum_{i=1}^{n} \sum_{j=-\ell}^{\ell} \sum_{t=\max(1,j+1)}^{\min(T,T+j)} \tilde{x}_{2it}\bigl(\tilde{y}_{i,t-j} - \tilde{x}_{2i,t-j}'\widehat{\theta}\bigr).$$

The one-step bias-corrected estimates of the common parameter θ and the average of the individual parameter $\alpha := E[\alpha_i]$ are

$$\widehat{\theta}^{\mathrm{BC}} = \widehat{\theta} - \widehat{\mathcal{B}}(\widehat{\theta})/T, \qquad \widehat{\alpha}^{\mathrm{BC}} = n^{-1}\sum_{i=1}^{n} \widehat{\alpha}_i\bigl(\widehat{\theta}^{\mathrm{BC}}\bigr).$$

The iterated bias correction estimator can be derived analytically by solving

$$\widehat{\theta}^{\mathrm{IBC}} = \widehat{\theta} - \widehat{\mathcal{B}}\bigl(\widehat{\theta}^{\mathrm{IBC}}\bigr)/T,$$

which has the closed-form solution

$$\widehat{\theta}^{\mathrm{IBC}} = \left[I_{d_\theta} + (d_g - d_\alpha)\bigl(\widehat{J}_s^W\bigr)^{-1} \sum_{i=1}^{n} \sum_{j=-\ell}^{\ell} \sum_{t=\max(1,j+1)}^{\min(T,T+j)} \tilde{x}_{2it}\tilde{x}_{2i,t-j}'/\bigl(nT^2\bigr)\right]^{-1}$$
$$\times \left[\widehat{\theta} + (d_g - d_\alpha)\bigl(\widehat{J}_s^W\bigr)^{-1} \sum_{i=1}^{n} \sum_{j=-\ell}^{\ell} \sum_{t=\max(1,j+1)}^{\min(T,T+j)} \tilde{x}_{2it}\tilde{y}_{i,t-j}/\bigl(nT^2\bigr)\right].$$

The score bias correction is the same as the iterated correction because the first-order conditions are linear in θ. □

The bias correction methods described above yield normal asymptotic distributions centered at the true parameter value for panels where n and T grow at the same rate with the sample size. This result is formally stated in Theorem 5, which establishes that all the methods are asymptotically equivalent, up to first order.

THEOREM 5 (Limit Distribution of Bias-Corrected FE-GMM). *Assume that* $\sqrt{nT}(\widehat{B}_s(\bar{\theta}) - B_s)/T \xrightarrow{p} 0$ *and* $\sqrt{nT}(\widehat{J}_s(\bar{\theta}) - J_s)/T \xrightarrow{p} 0$ *for some* $\bar{\theta} = \theta_0 + O_P((nT)^{-1/2})$. *Under Conditions 1–6, for* $C \in \{\text{BC}, \text{SBC}, \text{IBC}\}$,

$$\sqrt{nT}(\widehat{\theta}^C - \theta_0) \xrightarrow{d} N(0, J_s^{-1}), \tag{5.4}$$

where $\widehat{\theta}^{\text{BC}}$, $\widehat{\theta}^{\text{IBC}}$, *and* $\widehat{\theta}^{\text{SBC}}$ *are defined in* (5.1), (5.2), *and* (5.3), *and* $J_s = \bar{E}[G'_{\theta_i} P_{\alpha_i} G_{\theta_i}]$.

The convergence condition for the estimators of B_s and J_s holds for sample analogs evaluated at the initial FE-GMM one-step or two-step estimators if the trimming sequence is chosen such that $\ell \to \infty$ and $\ell/T \to 0$ as $T \to \infty$. Theorem 5 also shows that all the bias-corrected estimators considered are first-order asymptotically efficient, since their variances achieve the semiparametric efficiency bound for the common parameters in this model; see Chamberlain (1992).

The following corollaries give bias-corrected estimators for averages of the data and individual effects and for moments of the individual effects, together with the limit distributions of these estimators and consistent estimators of their asymptotic variances. To construct the corrections, we use bias-corrected estimators of the common parameter. The corollaries then follow from Lemma 2 and Theorem 5 by the delta method. We use the same notation as in the estimation of the bias of the common parameters above to denote the estimators of the components of the bias and variance.

COROLLARY 3 (Bias Correction for Fixed Effects Averages). *Let* $\zeta(z;\theta,\alpha_i)$ *be a twice continuously differentiable function in its second and third argument such that* $\inf_i \text{Var}[\zeta(z_{it})] > 0$, $\bar{E}E[\zeta(z_{it})^2] < \infty$, $\bar{E}E[\zeta_\alpha(z_{it})^2] < \infty$, *and* $\bar{E}E|\zeta_\theta(z_{it})|^2 < \infty$. *For* $C \in \{\text{BC}, \text{SBC}, \text{IBC}\}$, *let* $\widehat{\zeta}^C = \widehat{\zeta}(\widehat{\theta}^C) - \widehat{B}_\zeta(\widehat{\theta}^C)/T$, *where*

$$\widehat{B}_\zeta(\theta) = \widehat{\bar{E}}\Bigg[\sum_{j=0}^{\ell} \frac{1}{T} \sum_{t=j+1}^{T} \widehat{\zeta}_{\alpha_{it}}(\theta)' \widehat{\tilde{\psi}}_{\alpha_{i,t-j}}(\theta) + \widehat{\zeta}_{\alpha_i}(\theta)' \widehat{B}_{\alpha_i}(\theta) + \sum_{j=1}^{d_\alpha} \widehat{\zeta}_{\alpha\alpha_{i,j}}(\theta)' \widehat{\Sigma}_{\alpha_i}(\theta)/2\Bigg],$$

where ℓ *is a positive bandwidth such that* $\ell \to \infty$ *and* $\ell/T \to 0$ *as* $T \to \infty$. *Then, under the conditions of Theorem 5,*

$$r_{nT}(\widehat{\zeta}^C - \zeta) \xrightarrow{d} N(0, V_\zeta),$$

where r_{nT}, ζ, *and* V_ζ *are defined in Corollary 1. Also, for any* $\bar{\theta} = \theta_0 + O_P((nT)^{-1/2})$ *and* $\bar{\zeta} = \zeta + O_P(r_{nT}^{-1})$,

$$\widehat{V}_\zeta = \frac{r_{nT}^2}{nT}\widehat{\bar{E}}\{\widehat{E}[\widehat{\zeta}_{\alpha_{it}}(\bar{\theta})'\widehat{\Sigma}_{\alpha_i}(\bar{\theta})\widehat{\zeta}_{\alpha_{it}}(\bar{\theta}) + \widehat{\zeta}_{\theta_{it}}(\bar{\theta})'\widehat{J}_s(\bar{\theta})^{-1}\widehat{\zeta}_{\theta_{it}}(\bar{\theta})] + T(\widehat{E}[\widehat{\zeta}_{it}(\bar{\theta}) - \bar{\zeta}])^2\}$$

is a consistent estimator for V_ζ.

Corollary 4 (Bias Correction for Smooth Functions of Individual Effects). *Let* $\mu(\alpha_i)$ *be a twice differentiable function such that* $\bar{E}[\mu(\alpha_{i0})^2] < \infty$ *and* $\bar{E}|\mu_\alpha(\alpha_{i0})|^2 < \infty$. *For* $C \in \{\text{BC}, \text{SBC}, \text{IBC}\}$, *let* $\widehat{\mu}^C = \widehat{\bar{E}}[\widehat{\mu}_i(\widehat{\theta}^C)] - \widehat{B}_\mu(\widehat{\theta}^C)/T$, *where* $\widehat{\mu}_i(\theta) = \mu(\widehat{\alpha}_i(\theta))$ *and* $\widehat{B}_\mu(\theta) = \widehat{\bar{E}}[\widehat{\mu}_{\alpha_i}(\theta)'\widehat{B}_{\alpha_i}(\theta) + \sum_{j=1}^{d_\alpha}\widehat{\mu}_{\alpha\alpha_{i,j}}(\theta)'\widehat{\Sigma}_{\alpha_i}(\theta)/2]$. *Then, under the conditions of Theorem 5,*

$$\sqrt{n}(\widehat{\mu}^C - \mu) \xrightarrow{d} N(0, V_\mu),$$

where $\mu = \bar{E}[\mu(\alpha_{i0})]$ *and* $V_\mu = \bar{E}[(\mu(\alpha_{i0}) - \mu)^2]$. *Also, for any* $\bar{\theta} = \theta_0 + O_P((nT)^{-1/2})$ *and* $\bar{\mu} = \mu + O_P(n^{-1/2})$,

$$\widehat{V}_\mu = \widehat{\bar{E}}[\{\widehat{\mu}_i(\bar{\theta}) - \bar{\mu}\}^2 + \widehat{\mu}_{\alpha_i}(\bar{\theta})'\widehat{\Sigma}_{\alpha_i}(\bar{\theta})\widehat{\mu}_{\alpha_i}(\bar{\theta})/T] \tag{5.5}$$

is a consistent estimator for V_μ. *The second term in (5.5) is included to improve the finite-sample properties of the estimator in short panels.*

6. Empirical example

We illustrate the new estimators with an empirical example based on the classical cigarette demand study of Becker, Grossman, and Murphy (1994) (BGM hereafter). Cigarettes are addictive goods. To account for this addictive nature, early cigarette demand studies included lags of consumption as explanatory variables (e.g., Baltagi and Levin (1986)). This approach, however, ignores that rational or forward-looking consumers take into account the effect of today's consumption decision on future consumption decisions. Becker and Murphy (1988) developed a model of rational addiction where expected changes in future prices affect the current consumption. BGM empirically tested this model using a linear structural demand function based on quadratic utility assumptions. The demand function includes both future and past consumption as determinants of current demand, and the future price affects the current demand only through future consumption. They found that the effect of future consumption on current consumption is significant, which they took as evidence in favor of the rational model.

Most of the empirical studies in this literature use yearly state-level panel data sets. They include fixed effects to control for additive heterogeneity at the state level, and use leads and lags of cigarette prices and taxes as instruments for leads and lags of consumption. These studies, however, do not consider possible nonadditive heterogeneity in price effects across states. There are multiple reasons why there may be heterogeneity in the price effect across states correlated with the price level. First, the considerable differences in income, industrial, ethnic, and religious composition at the interstate level can translate into different tastes and policies toward cigarettes. Second, from the perspective of the theoretical model developed by Becker and Murphy (1988), the price effect is a function of the marginal utility of wealth that varies across states and depends on cigarette prices. If the price effect is heterogenous and correlated with the price level, a fixed-coefficient specification may produce substantial bias in estimating the average elasticity of cigarette consumption because the between variation of price is much larger than the within variation. Wangen (2004) gave additional theoretical reasons against a fixed-coefficient specification for the demand function in this application.

For the demand function, we consider the linear specification

$$C_{it} = \alpha_{0i} + \alpha_{1i} P_{it} + \theta_1 C_{i,t-1} + \theta_2 C_{i,t+1} + X_{it}' \delta + \varepsilon_{it}, \tag{6.1}$$

where C_{it} is cigarette consumption in state i at time t measured by per capita sales in packs, α_{0i} is an additive state effect, α_{1i} is a state-specific price coefficient, P_{it} is the price in 1982–1984 dollars, and X_{it} is a vector of covariates that includes income, various measures of incentive for smuggling across states, and year dummies. We estimate the model parameters using OLS and IV methods with both a fixed coefficient for price and a random coefficient for price. The data set, consisting of an unbalanced panel of 50 U.S. states and the district of Columbia over the years 1957 to 1994, is the same as in Fenn, Antonovitz, and Schroeter (2001). The set of instruments for $C_{i,t-1}$ and $C_{i,t+1}$ in the IV estimators is the same as in specification 3 of BGM and includes X_{it}, P_{it}, $P_{i,t-1}$, $P_{i,t+1}$, Tax_{it}, $\text{Tax}_{i,t-1}$, and $\text{Tax}_{i,t+1}$, where Tax_{it} is the state excise tax for cigarettes in 1982–1984 dollars.

Table 1 reports estimates of coefficients and demand elasticities. We focus on the coefficients of the key variables, namely P_{it}, $C_{i,t-1}$, and $C_{i,t+1}$. Throughout the table, FC refers to the fixed-coefficient specification with $\alpha_{1i} = \alpha_1$ and RC refers to the random-coefficient specification in equation (6.1). BC and IBC refer to estimates after bias correction and iterated bias correction, respectively. Demand elasticities are calculated using the expressions in Appendix A of BGM. They are functions of C_{it},P_{it}, α_{1i}, θ_1, and θ_2, linear in α_{1i}. For random-coefficient estimators, we report the mean of individual elasticities, that is,

$$\widehat{\zeta}_h = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \zeta_h(z_{it}; \widehat{\theta}, \widehat{\alpha}_i),$$

where $\zeta_h(z_{it}; \theta, \alpha_i) = \partial \log C_{it(h)} / \partial \log P_{it(h)}$ are price elasticities at different time horizons h. Standard errors for the elasticities are obtained by the delta method as described in Corollary 3. For bias-corrected RC estimators, the standard errors use bias-corrected estimates of θ and α_i.

As BGM did, we find that OLS estimates substantially differ from their IV counterparts. IV-FC underestimates the elasticities relative to IV-RC. For example, the long-run elasticity estimate is -0.70 with IV-FC, whereas it is -0.88 with IV-RC. This difference is also pronounced for short-run elasticities, where the IV-RC estimates are more than 25% larger than the IV-FC estimates. We observe the same pattern throughout the table for every elasticity. The bias comes from both the estimation of the common parameter θ_2

TABLE 1. Estimates of rational addiction model for cigarette demand.

			OLS-RC			IV-RC		
	OLS-FC	IV-FC	NBC	BC	IBC	NBC	BC	IBC
Coefficients								
(Mean) P_t	−9.58	−34.10	−13.49	−13.58	−13.26	−36.39	−31.26	−31.26
	(1.86)	(4.10)	(3.55)	(3.55)	(3.55)	(4.85)	(4.62)	(4.64)
(Std. dev.) P_t			4.35	4.22	4.07	12.86	10.45	10.60
			(0.98)	(1.02)	(1.03)	(2.35)	(2.13)	(2.15)
C_{t-1}	0.49	0.45	0.48	0.48	0.48	0.44	0.44	0.45
	(0.01)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
C_{t+1}	0.44	0.17	0.44	0.43	0.44	0.23	0.29	0.27
	(0.01)	(0.07)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Price elasticities								
Long run	−1.05	−0.70	−1.30	−1.31	−1.28	−0.88	−0.91	−0.90
	(0.24)	(0.12)	(0.28)	(0.28)	(0.28)	(0.09)	(0.10)	(0.10)
Own price	−0.20	−0.32	−0.27	−0.27	−0.27	−0.38	−0.35	−0.35
(anticipated)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)
Own price	−0.11	−0.29	−0.15	−0.16	−0.15	−0.33	−0.29	−0.29
(unanticipated)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Future price	−0.07	−0.05	−0.10	−0.10	−0.09	−0.09	−0.10	−0.09
(unanticipated)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Past price	−0.08	−0.14	−0.11	−0.11	−0.10	−0.16	−0.15	−0.15
(unanticipated)	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Short run	−0.30	−0.35	−0.41	−0.41	−0.40	−0.44	−0.44	−0.43
	(0.05)	(0.06)	(0.12)	(0.12)	(0.12)	(0.06)	(0.06)	(0.06)

Note: RC and FC refer to the random and the fixed-coefficient model. NBC, BC, and IBC refer to no bias correction, bias correction, and iterated bias correction estimates. Standard errors are given in parentheses.

and the mean of the individual-specific parameter $E[\alpha_{1i}]$. The bias corrections increase the coefficient of future consumption $C_{i,t+1}$ and reduce the absolute value of the mean of the price coefficient. Moreover, they have a significant impact on the estimator of dispersion of the price coefficient. The uncorrected estimates of the standard deviation are more than 20% larger than their bias-corrected counterparts. In the Appendix, we show through a Monte Carlo experiment calibrated to this empirical example that the bias is generally large for dispersion parameters and the bias corrections are effective in reducing this bias. As a consequence of shrinking the estimates of the dispersion of α_{1i}, we obtain smaller standard errors for the estimates of $E[\alpha_{1i}]$ throughout the table. In the Monte Carlo experiment, we also find that this correction in the standard errors provides improved inference.

7. Conclusion

This paper introduces a new class of fixed effects GMM estimators for panel data models with unrestricted nonadditive heterogeneity and endogenous regressors. Bias correction methods are developed because these estimators suffer from the incidental parameters problem. Other estimators based on moment conditions, like the class of generalized empirical likelihood (GEL) estimators, can be analyzed using a similar methodology. An attractive alternative framework for estimation and inference in random-coefficient models is a flexible Bayesian approach. It would be interesting to explore whether there are connections between moments of posterior distributions in the Bayesian approach and the fixed effects estimators considered in the paper. Another interesting extension would be to find bias reducing priors in the GMM framework similar to those characterized by Arellano and Bonhomme (2009) in the MLE framework. We leave these extensions to future research.

References

Alvarez, J. and M. Arellano (2003), "The time series and cross-section asymptotics of dynamic panel data estimators." *Econometrica*, 71, 1121–1159. [455]

Angrist, J. D. (2004), "Treatment effect heterogeneity in theory and practice." *The Economic Journal*, 114 (494), C52–C83. [454]

Angrist, J. D., K. Graddy, and G. W. Imbens (2000), "The interpretation of instrumental variables estimators in simultaneous equation models with an application to the demand of fish." *Review of Economic Studies*, 67, 499–527. [454, 458]

Angrist, J. D. and J. Hahn (2004), "When to control for covariates? Panel asymptotics for estimates of treatment effects." *Review of Economics and Statistics*, 86 (1), 58–72. [457]

Angrist, J. D. and G. W. Imbens (1995), "Two-stage least squares estimation of average causal effects in models with variable treatment intensity." *Journal of the American Statistical Association*, 90, 431–442. [458]

Angrist, J. D. and A. B. Krueger (1999), "Empirical strategies in labor economics." In *Handbook of Labor Economics*, Vol. 3 (O. Ashenfelter and D. Card, eds.), North-Holland, Amsterdam. [458]

Arellano, M. and S. Bonhomme (2009), "Robust priors in nonlinear panel data models." *Econometrica*, 77, 489–536. [478]

Arellano, M. and S. Bonhomme (2012), "Identifying distributional characteristics in random coefficients panel data model." *Review of Economic Studies*, 79 (3), 987–1020. [456]

Arellano, M. and J. Hahn (2006), "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects." Report, CEMFI. [455]

Arellano, M. and J. Hahn (2007), "Understating bias in nonlinear panel models: Some recent developments." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, Vol. 3 (R. Blundell, W. K. Newey, and T. Persson, eds.), Cambridge University Press, Cambridge. [455]

Bai, J. (2009), "Panel data models with interactive fixed effects." *Econometrica*, 77 (4), 1229–1279. [455]

Baltagi, B. H. and D. Levin (1986), "Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered." *Review of Economics and Statistics*, 68, 148–155. [475]

Becker, G. S., M. Grossman, and K. M. Murphy (1994), "An empirical analysis of cigarette addiction." *American Economic Review*, 84, 396–418. [453, 455, 475]

Becker, G. S. and K. M. Murphy (1988), "A theory of rational addiction." *Journal of Political Economy*, 96, 675–700. [475, 476]

Bester, A. and C. Hansen (2008), "A penalty function approach to bias reduction in nonlinear panel models with fixed effects." *Journal of Business & Economic Statistics*, 27 (2), 131–148. [455]

Buse, A. (1992), "The bias of instrumental variables estimators." *Econometrica*, 60, 173–180. [459]

Chamberlain, G. (1992), "Efficiency bounds for semiparametric regression." *Econometrica*, 60, 567–596. [474]

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. K. Newey (2013), "Average and quantile effects in nonseparable panel models." Unpublished manuscript, MIT. [456]

Dhaene, G. and K. Jochmans (2010), "Split-panel jackknife estimation of fixed effects models." Unpublished manuscript, K.U. Leuven. [471]

Fenn, A. J., F. Antonovitz, and J. R. Schroeter (2001), "Cigarettes and addiction information: New evidence in support of the rational addiction model." *Economics Letters*, 72, 39–45. [476]

Fernández-Val, I. (2005), *Three Essays on Nonlinear Panel Data Models and Quantile Regression Analysis*. Ph.D. dissertation, Massachusetts Institute of Technology. DSpace@MIT. [453]

Graham, B. S. and J. L. Powell (2012), "Identification and estimation of average partial effects in 'irregular' correlated random coefficient panel data models." *Econometrica*, 80 (5), 2105–2152. [456]

Hahn, J. and G. Kuersteiner (2002), "Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large." *Econometrica*, 70, 1639–1657. [455]

Hahn, J. and G. Kuersteiner (2011), "Bias reduction for dynamic nonlinear panel models with fixed effects." *Econometric Theory*, 27, 1152–1191. [455, 464, 465, 472]

Hahn, J., G. Kuersteiner, and W. Newey (2004), "Higher order properties of bootstrap and jackknife bias corrections." Unpublished manuscript. [471]

Hahn, J. and W. Newey (2004), "Jackknife and analytical bias reduction for nonlinear panel models." *Econometrica*, 72, 1295–1319. [455, 467, 471, 473]

Hansen, L. P. (1982), "Large sample properties of generalized method of moments estimators." *Econometrica*, 50, 1029–1054. [462]

Heckman, J. and E. Vytlacil (2000), "Instrumental variables methods for the correlated random coefficient model." *Journal of Human Resources*, 33 (4), 974–987. [454]

Hsiao, C. and M. H. Pesaran (2004), "Random coefficient panel data models." Report, University of Southern California. [454]

Kelejian, H. H. (1974), "Random parameters in a simultaneous equation framework: Identification and estimation." *Econometrica*, 42 (3), 517–528. [454]

Kiviet, J. F. (1995), "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models." *Journal of Econometrics*, 68 (1), 53–78. [455]

Lancaster, T. (2002), "Orthogonal parameters and panel data." *Review of Economic Studies*, 69, 647–666. [455]

MacKinnon, J. G. and A. A. Smith (1998), "Approximate bias correction in econometrics." *Journal of Econometrics*, 85, 205–230. [472]

Murtazashvili, I. and J. M. Wooldridge (2005), "Fixed effects instrumental variables estimation in correlated random coefficient panel data models." Unpublished manuscript, Michigan State University. [458]

Nagar, A. L. (1959), "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations." *Econometrica*, 27, 575–595. [459]

Newey, W. K. and D. McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics*, Vol. 4 (R. F. Engle and D. L. McFadden, eds.), Elsevier Science, Amsterdam. [464, 465]

Newey, W. K. and R. Smith (2004), "Higher order properties of GMM and generalized empirical likelihood estimators." *Econometrica*, 72, 219–255. [454, 455, 462, 469]

Neyman, J. and E. L. Scott (1948), "Consistent estimates based on partially consistent observations." *Econometrica*, 16, 1–32. [459]

Phillips, P. C. B. and H. R. Moon (1999), "Linear regression limit theory for nonstationary panel data." *Econometrica* 67, 1057–1111. [455]

Rilstone, P., V. K. Srivastava, and A. Ullah (1996), "The second-order bias and mean squared error of nonlinear estimators." *Journal of Econometrics*, 75, 369–395. [459]

Roy, A. (1951), "Some thoughts on the distribution of earnings." *Oxford Economic Papers*, 3, 135–146. [454]

Wangen, K. R. (2004), "Some fundamental problems in Becker, Grossman and Murphy's implementation of rational addiction theory." Discussion Papers 375, Research Department of Statistics Norway. [476]

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge. [457]

Wooldridge, J. M. (2005), "Fixed effects and related estimators for correlated random-coefficient and treatment-effect panel data models." *Review of Economics and Statistics*, 87 (2), 385–390. [458]

Woutersen, T. M. (2002), "Robustness against incidental parameters." Unpublished manuscript, University of Western Ontario. [455]

Yitzhaki, S. (1996), "On using linear regressions in welfare economics." *Journal of Business & Economic Statistics*, 14, 478–486. [454, 458]

Submitted April, 2010. Final version accepted January, 2013.