FIRST-DIFFERENCING IN PANEL DATA MODELS WITH INCIDENTAL FUNCTIONS

KOEN JOCHMANS*

CORE AND BROWN UNIVERSITY

This version: October, 20 2010

ABSTRACT

I discuss the fixed-effect estimation of panel data models with time-varying excess heterogeneity across cross-sectional units. These latent components are not given a parametric form. A modification to traditional first-differencing is motivated which, asymptotically, removes the permanent unobserved heterogeneity from the differenced model. Conventional estimation techniques can then be readily applied. Distribution theory for a kernel-weighted GMM estimator under large-n and fixed-T asymptotics is developed. The estimator is put to work in a series of numerical experiments to static and dynamic models.

JEL CLASSIFICATION: C13, C14, C33

KEYWORDS: dynamic panel data, GMM, incidental functions, local first-differencing, time-varying fixed effects, nonparametric heterogeneity.

*Address: Center for Operations Research and Econometrics, U.C. Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium. Tel. +32 10 474 329; E-mail: koen.jochmans@uclouvain.be. Or: Department of Economics, Brown University, Box B, Providence, RI 02912, U.S.A. Tel. +1 401 863 2970; E-mail: koen_jochmans@brown.edu. This note was inspired by the *Journal of Applied Econometrics* annual lecture of Manuel Arellano at the 16th International Conference on Panel Data in Amsterdam [July 2010], entitled 'Nonparametric random coefficients in panel data models'. The first version of this document dates from July 2010.

I INTRODUCTION

The use of longitudinal data in microeconometric studies is commonly motivated by the potential to capture excess heterogeneity between units. The first major rationale for this is to obtain parameter estimates that are free from bias induced by unobserved cross-sectional heterogeneity. A second reason for their employment is the desire to decompose the variance of latent components into structural and transitory contributions. The former incentive is typically related to taking a fixed-effect view on the processes that drive the outcome of interest in the sense of leaving the distribution of unobserved unit-specific factors unspecified. The error-component formulation or random-effect approach is commonly associated with stronger restrictions on the distribution of the permanent latent component.

So far, in the spirit of the seminal work of Mundlack (1961, 1978), the most common specification adds unit-specific intercept terms to a model that is otherwise homogeneous in the effect of covariates on outcomes. This captures time-constant and non-interactive unobserved components. Nevertheless, there may be good reason to assume that the heterogeneity among processes is wider spread. Browning, Ejrnæs, and Alvarez (2009) and Arellano and Bonhomme (2010), for example, provide ellaborate discussions on this and report empirical results that support their claims. In light of this, there has been renewed attention toward identifying the distribution of random coefficients in linear panel data models.

In this paper, I complement such a strategy by considering what can be learned about common parameters in, possibly dynamic, panel data models with nonparametrically-specified excess heterogeneity. Here, the fixed-effect paradigm—that is, circumventing the estimation of the model's idiosyncratic components—allows for a very rich pattern of unobserved heterogeneity in a manner that leaves its distribution essentially unrestricted. The heterogenous effects may enter non-additively and can be time-varying. This leads to a framework in which the conventional incidental parameters are replaced by incidental functions, and traditional first-differencing has to make way for local first-differencing.

The estimator I describe below relates to theory on semi- and nonparametric cross-sectional estimation derived by Gozalo and Linton (2000) and Lewbel (2007), and to kernel-weighted pairwise-differencing estimation as first introduced by Powell (1987). However, observing multiple time-series realizations for the same units allows tackling heterogeneity in a manner that pairwise-differencing cannot. In a panel data context, the reference most closely related to what follows below is Honoré and Kyriazidou (2000), who proposed a kernel-weighted conditional-likelihood estimator for a dynamic logit model with time-varying covariates. In fact, this connection immediately suggests that the proposed approach to dealing with incidental functions can be generalized to all existing fixed-effect estimators, such as conditional maximum-likelihood and maximum score, for example. This is easy to see, but working out the details for the general case would come at the cost of reduced transparency.

This paper has four more sections. The first of these provides details on the model of interest and the local-differencing strategy, and states the GMM estimator that will be the focus of the analysis. The next section contains large-sample theory. The last section before concluding gives results from a series of Monte Carlo exercises.

II THE MODEL, LOCAL FIRST DIFFERENCING, AND A GMM ESTIMATOR

Suppose we are in possession of n independent sequences of T observations on the variables (y, x, z, v). Denote the tth observation for unit $i = 1, \ldots, n$ by $(y_{it}, x_{it}, z_{it}, v_{it})$ and let $\ell(a)$ be the dimension of the vector a. Assume that the observations on the scalar random variable y were generated through a model of the form

$$y_{it} = x_{it}'\theta_0 + \xi_{it}, \qquad \xi_{it} = \vartheta_i(v_{it}) + \varepsilon_{it}. \tag{2.1}$$

Here, ε_{it} is a zero mean random disturbance, θ_0 is the true value of the parameter vector of interest, and the unit-specific functions $\vartheta_i : \mathcal{R}^{\ell(v)} \rightarrow \mathcal{R}$ capture the, potentially heterogeneous, impact of the covariates v on y. These functions are assumed to be smooth in their arguments, but are otherwise unmodelled and allowed to be random across i. Remaining loyal to the fixed-effect modelling approach, they are draws from an unknown probability distribution that may depend on the realization of the covariates and, possibly, initial conditions, but not on ε. Feedback from y_t to future x_t is allowed for, accommodating dynamic models with predetermined regressors. More generally, x is allowed to be endogenous in the cross-sectional sense of being contemporaneously correlated with ε. When strict exogeneity is lacking, z serves as instrumental variables. No restriction is put on the statistical relationship between v and (x, ε) over time.

The fact that $\vartheta_i(v)$ may be random after conditioning on a realization of its argument distinguishes (2.1) from a panel data version of a nonparametric functional-coefficient model as considered by Hastie and Tibshirani (1993). It is also different from the random-coefficient model of Chamberlain (1992) and Arellano and Bonhomme (2010), where $\vartheta_i(v)$ is a linear combination of exogenous v and unit-specific coefficients. Here, the ϑ_i are unspecified incidental *functions* of covariates, which may be predetermined or endogenous. Thus, (2.1) generalizes the standard linear fixed-effect model by allowing for a much richer pattern of unobserved unit-specific factors; notice that this benchmark model is obtained on assuming $\vartheta_i(v)$ to be constant in v.

To construct an estimator of θ_0, start by stacking observations over time to obtain $y_i \equiv (y_{i1}, \ldots, y_{iT})'$, $X_i \equiv (x_{i1}, \ldots, x_{iT})'$, $V_i \equiv (v_{i1}, \ldots, v_{iT})'$, $\varepsilon_i \equiv (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, and $\xi_i \equiv (\xi_{i1}, \ldots, \xi_{iT})'$; let $\vartheta_i(V_i) \equiv (\vartheta_i(v_{i1}), \ldots, \vartheta_i(v_{iT}))'$. Write D for the $(T-1)\times T$ matrix that performs first-differencing on the above arrays. For example, $\mathrm{D}y_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})'$, where $\Delta \equiv 1 - \mathrm{L}$ and L denotes the lag operator. Then,

$$y_i = X_i\theta_0 + \xi_i, \qquad \xi_i = \vartheta_i(V_i) + \varepsilon_i, \tag{2.2}$$

and $\mathrm{D}y_i = \mathrm{D}X_i\theta_0 + \mathrm{D}\xi_i$ for $\mathrm{D}\xi_i = \mathrm{D}\vartheta_i(V_i) + \mathrm{D}\varepsilon_i$. Recall that, when $\vartheta_i(v)$ is constant, $\mathrm{D}\vartheta_i(V_i) = 0$, and an efficient GMM estimator of θ_0 that is based on data in first-differences minimizes the quadratic form

$$\Big[\frac{1}{n}\sum_{i=1}^{n} Z_i'\mathrm{D}(y_i - X_i\theta)\Big]' G \Big[\frac{1}{n}\sum_{i=1}^{n} Z_i'\mathrm{D}(y_i - X_i\theta)\Big] \tag{2.3}$$

for a well-chosen weight matrix G; see Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991). Throughout, Z_i is an instrument matrix of conformable dimension that contains $z_{i2}, \ldots, z_{iT}$. The latter can be lagged levels of predetermined regressors, all time-series realizations of exogenous covariates, and, possibly, external instrumental variables. Of course, simple differencing does not remove $\vartheta_i(V_i)$ from (2.2) more generally, and $\mathscr{E}\big[Z'\mathrm{D}(y - X\theta_0)\big] \neq 0$ unless z_t and $\Delta\vartheta(v_t)$ are uncorrelated.[1] However, by smoothness of ϑ_i, $\Delta y_{it} = \Delta x_{it}'\theta_0 + \Delta\vartheta_i(v_{it}) + \Delta\varepsilon_{it} \rightarrow \Delta x_{it}'\theta_0 + \Delta\varepsilon_{it}$ as $\Delta v_{it} \rightarrow 0$. Let $\omega_{\mathrm{v}}(v) \equiv \delta(v - \mathrm{v})$ for $\delta : \mathcal{R}^{\ell(v)} \rightarrow \mathcal{R}$ a multivariate version of Dirac's delta.[2] Then, the conditional moment condition $\mathscr{E}\big[z_t \ \omega_0(\Delta v_t)(\Delta y_t - \Delta x_t'\theta_0)\big] = 0$, which is free of incidental functions, holds for $t = 2, \ldots, T$. This suggest constructing an estimator based on an empirical counterpart of these conditional moment conditions. A fruitful way to proceed is to replace $\omega_0(\Delta v_t)$ by a smoother such as a kernel weight, and to make the appropriate changes to (2.3).

---

[1] As interest here lies in micropanels, throughout, the maintained asymptotic scheme will be that of large n and fixed T. Accordingly, expectations are taken with respect to the cross-sectional distribution of the data.

[2] Dirac's delta is defined as follows: if $a \neq 0$, then $\delta(a) = 0$; if $a = 0$, then $\delta(a) = \infty$. As a measure it has the property that $\int g(a)\ \delta(a)\ \mathrm{d}a = g(0)$ for a continuous function g. I prefer its use here over the usual notation for conditional expectations because explicitly conditioning on all time-series realizations of covariates and instruments on top of period-specific first differences would impose a heavy notational burden.

To do so, partition the vector v as $(v^{(c)}, v^{(d)})$, where $v^{(c)}$ and $v^{(d)}$ refer to the continuously-distributed components and the discrete components of v, respectively. Let $\ell \equiv \ell(v^{(c)})$, let $\mathrm{k} : \mathcal{R}^\ell \to \mathcal{R}$ be a kernel with associated bandwidth σ_n, and write $1(a)$ for the indicator function for the event a. A kernel weight takes the form

$$w_{it} \equiv \frac{1}{\sigma_n^\ell} \mathrm{k}\Big(\frac{\Delta v_{it}^{(c)}}{\sigma_n}\Big)\ 1(\Delta v_{it}^{(d)} = 0).$$

Form the diagonal matrices $W_i \equiv \operatorname{diag}(w_{i2}, \ldots, w_{iT})$ and $\Omega_{\mathrm{v}} \equiv \operatorname{diag}(\omega_{\mathrm{v}}(v_2), \ldots, \omega_{\mathrm{v}}(v_T))$. A local GMM estimator based on the moment conditions

$$\mathscr{E}\big[Z'\Omega_0 \mathrm{D}(y - X\theta_0)\big] = \mathscr{E}\big[Z'\Omega_0 \mathrm{D}\varepsilon\big] = 0 \tag{2.4}$$

is obtained by introducing W_i in (2.3). Because the weights do not distort the linearity of this objective function, the local estimator is available in closed form. Moreover, it is given by

$$\theta_n \equiv \big[S'_{ZX} G S_{ZX}\big]^{-1} \big[S'_{ZX} G S_{Zy}\big] \tag{2.5}$$

for the matrices

$$S_{ZX} \equiv \frac{1}{n}\sum_{i=1}^n Z_i' W_i \mathrm{D} X_i \quad \text{and} \quad S_{Zy} \equiv \frac{1}{n}\sum_{i=1}^n Z_i' W_i \mathrm{D} y_i.$$

Thus, computing θ_n from the data boils down to first assigning a weight to each first-differenced observation and then proceeding as one would do with the conventional GMM estimator of choice. Under the regularity conditions provided below, this approach is asymptotically equivalent to using only first-differenced observations for which Δv_{it} lies in a shrinking neighborhood of zero. Observe that (2.5) reduces to a weighted least-squares estimator on setting $Z_i = \mathrm{D}X_i$.

One attractive consequence of working with data in first differences is that it is straightforward to deal with unbalanced panel data. Nevertheless, estimation based on weighted versions of orthogonal deviations (Arellano and Bover, 1995) or deviations from within-group means is also possible. To illustrate, let

$$\dot{w}_{it}^s \equiv \frac{1}{\sigma_n^\ell} \mathrm{k}\Big(\frac{v_{it}^{(c)} - v_{is}^{(c)}}{\sigma_n}\Big)\ 1(v_{it}^{(d)} = v_{is}^{(d)}),$$

choose k to be symmetric, and collect the $\dot{w}_{it}^s$ in the symmetric $T \times T$ matrix $\dot{W}_i$, the (t,s)th element of this matrix being $\dot{w}_{it}^s$. A local GMM estimator that is based on forward orthogonal deviations follows on replacing S_{XZ} and S_{Xy} above by the matrices

$n^{-1}\sum_{i=1}^n Z_i'(\mathrm{DD}')^{-1/2}\mathrm{D}\dot{W}_i X_i$ and $n^{-1}\sum_{i=1}^n Z_i'(\mathrm{DD}')^{-1/2}\mathrm{D}\dot{W}_i y_i$, respectively. A weighted within-group estimator, on the other hand, would take the form $\arg\min_\theta n^{-1}\sum_{i=1}^n (y_i - X_i\theta)'\dot{W}_i\mathrm{D}'(\mathrm{DD}')^{-1}\mathrm{D}\dot{W}_i(y_i - X_i\theta)$. The matrix $\mathrm{D}'(\mathrm{DD}')^{-1}\mathrm{D}$ may be recalled to be the traditional within-group operator.[3] Distribution theory for both these estimators can be derived in an analogous fashion as for θ_n. However, simulation results suggest that neither outperforms its first-differenced counterpart in small samples. The reason for this is the additional noise induced by taking linear combinations of kernel-weighted quantities.

III Large-sample behavior of the GMM estimator

Let us now consider the behavior of θ_n as n grows large. For vectors a and b for which $\ell(a) = \ell(b)$, let $\|a\|$ denote the Euclidean norm, let $|a| \equiv \sum_{j=1}^{\ell(a)} a^{(j)}$, and let $b^a \equiv \sum_{j=1}^{\ell(a)} (b^{(j)})^{a^{(j)}}$ throughout. The notation $\|\cdot\|$ will also be used to indicate the matrix norm.

Start by imposing the following conditions.

Assumption 1 (Regularities)**.** *The data array $\{y_i, x_i, z_i, v_i\}_{i=1}^n$ is a random sample. For $t = 2, \ldots, T$, $\mathscr{E}[\omega_0(\Delta v_t)\|z_t\Delta\varepsilon_t\|]$ is finite. For* v *in a neighborhood of zero, $\mathscr{E}[Z'\Omega_\mathrm{v}\mathrm{D}X]$ has full column rank. The moment condition $\mathscr{E}[Z'\Omega_0\mathrm{D}\varepsilon] = 0$ holds.*

Assumption 1 states the sampling scheme and contains conventional requirements for linear GMM estimators. Moreover, it provides a mild dominance condition on the vector of population moments and ensures identification of θ_0 by imposing a rank condition and postulating instrument validity.

The next assumption demands smoothness from the conditional moments that are being approximated by the kernel weighting and from the conditional densities of $\Delta v_t^{(c)}$ given realizations of $\Delta v_t^{(d)}$; refer to these densities by f_t and to their respective supports by $\mathcal{V}_t$.

Assumption 2 (Smoothness)**.** *For $t = 2, \ldots, T$, $f_t(\mathrm{v}^{(c)}|\mathrm{v}^{(d)})$ is bounded from above for all* v *in $\mathcal{V}_t$ and is strictly positive for* v *in a neighborhood of zero. In addition, $f_t(\mathrm{v}^{(c)}|\mathrm{v}^{(d)})$ and $\mathscr{E}[\omega_\mathrm{v}(\Delta v_t)z_t\Delta\xi_t]$ are k-times continuously differentiable in $\mathrm{v}^{(c)}$ for all* v *in a neighborhood of zero, where $k \geq 2$ is an integer.*

[3] Of course, the usual caveat in this interpretation applies when k is a higher-order kernel, as then some of the weights will have to be negative.

When combined with a well-behaved kernel and bandwidth, Assumption 2 implies that the bias induced by weighting observations by w_{it} disappears asymptotically. It could be relaxed slightly, for example by requiring the bounded away from zero condition on f_t to hold only for some but not all $t = 2, \ldots, T$, but this would needlessly cloud the exposition.

What is meant exactly by a well-behaved kernel and bandwidth is the topic of the third assumption.

Assumption 3 (Kernel weights)**.** *The kernel,* k*, is bounded on its support and of* k*th-order. Moreover,* $\int \mathrm{k}(\eta)\ \mathrm{d}\eta = 1$*,* $\int \eta^j\ \mathrm{k}(\eta)\ \mathrm{d}\eta = 0$ *for* $|j| = 1, \ldots, k-1$*, and* $\int \|\eta^j\|\ \|\mathrm{k}(\eta)\|\ \mathrm{d}\eta < \infty$ *for* $|j|$ *in* $\{0, k\}$*. The bandwidth,* σ_n*, is both non-negative and* $\mathscr{o}(1)$ *as* $n \to \infty$*, while* $\sqrt{n\sigma_n^\ell} \to \infty$ *and* $\sqrt{n\sigma_n^\ell}\sigma_n^k \to 0$*.*

The required degree of smoothness is increasing in the number of continuous components of v_{it}, ℓ. It is well known that, although they generally increase finite-sample bias, discrete covariates do not retard the speed of convergence of Nadaraya-Watson type kernel estimator of conditional mean functions, and the same conclusion may be drawn here.

It can be noted that an alternative to working with empirical cell probabilities for the discrete components of Δv_t would be to smooth over these variables as well. Such an approach may well be beneficial in small samples when the support of $\Delta v_t^{(d)}$ is large. Another refinement would be to work with regressor-specific bandwidths, that is, a vector-valued σ_n. I abstract away from both of these possibilities for fine tuning here as they do not affect the asymptotic behavior of θ_n, nor will it be possible at this stage to provide particular guidelines on how to get the most out of these additional degrees of freedom.

Assumptions 1–3 are more than enough to guarantee the consistency of θ_n for θ_0 as $n \to \infty$. To facilitate the exposition, let F_1 be a matrix consisting of $T-1$ vertically-stacked blocks. The $(t-1)$th such block is given by $\iota_{\ell(z_t)}\iota'_{\ell(\Delta x_t)}f_t(0|0)$ for ι_a a vector of ones of length a. On letting $\odot$ denote the Schur product,

$$S_{ZX} \xrightarrow{p} \Sigma_{ZX} \equiv \mathscr{E}[Z'\Omega_0 \mathrm{D}X] \odot F_1, \qquad S_{Zy} \xrightarrow{p} \Sigma_{Xy} \equiv \mathscr{E}[Z'\Omega_0 \mathrm{D}y] \odot F_1,$$

and $S_{Z\xi}(\theta) \equiv S_{Zy} - S_{ZX}\theta \xrightarrow{p} \Sigma_{Z\xi}(\theta) \equiv \Sigma_{Zy} - \Sigma_{ZX}\theta$. Write $S_{Z\xi}$ and $\Sigma_{Z\xi}$ for $S_{Z\xi}(\theta_0)$ and $\Sigma_{Z\xi}(\theta_0)$, respectively. The consistency result then reads as follows.

Theorem 1 (Consistency)**.** *Let Assumptions 1–3 hold. If* G *and* Γ *are positive-definite matrices so that* $G \xrightarrow{p} \Gamma$ *as* $n \to \infty$ *and* Γ *is non-stochastic, then* $\theta_n \xrightarrow{p} \theta_0$ *as* $n \to \infty$*.*

Proof. Recall that $S_{Z\xi} = n^{-1}\sum_{i=1}^{n} Z_i'W_i\mathrm{D}\xi_i$, so $\theta_n - \theta_0 = \left[S_{ZX}'GS_{ZX}\right]^{-1}\left[S_{ZX}'GS_{Z\xi}\right]$. Given that G converges to a positive-definite and non-stochastic matrix, it suffices to show that, as $n \to \infty$, (i) $S_{ZX} \xrightarrow{p} \Sigma_{ZX}$; and (ii) $S_{Z\xi} \xrightarrow{p} 0$.

To see that (i) holds, observe that standard kernel-smoothing arguments yield

$$\left\|\frac{1}{\sigma_n^{\ell}}\mathscr{E}\left[z_t\Delta x_t'\mathrm{k}\left(\frac{\Delta v_t^{(c)}}{\sigma_n}\right)1(\Delta v_t^{(d)}=0)\right]\right\| = \left\|\mathscr{E}\left[\omega_0(\Delta v_t)z_t\Delta x_t'\right]\ f_t(0|0)\right\| + \mathcal{O}(\sigma_n^{\ell})$$

for $t = 2,\ldots,T$. As $\mathscr{E}[S_{ZX}]$ is obtained on stacking such terms over t, it follows that $\|\mathscr{E}[S_{XZ}] - \Sigma_{XZ}\| = \mathcal{O}(\sigma_n^{\ell})$. This bias goes to zero because $\lim_{n\to\infty}\sigma_n^{\ell} = 0$, that is, $\overline{\mathscr{E}}[S_{ZX}] \equiv \lim_{n\to\infty}\mathscr{E}[S_{ZX}] = \Sigma_{ZX}$. Furthermore, by a standard law of large numbers, $\|S_{XZ} - \mathscr{E}[S_{XZ}]\| = o_p(1)$. Statement (i) follows.

By the same reasoning, $\|S_{Z\xi} - \Sigma_{Z\xi}\| = o_p(1)$. Instrument validity implies that $\Sigma_{Z\xi} = 0$. The proof is complete. □

Establishing asymptotic normality of θ_n around θ_0 requires the bias induced by kernel weighting to go to zero at a sufficiently fast rate, and the existence of higher-order conditional moments. For the former requirement to be satisfied, the current assumptions suffice. The latter is dealt with now. Let f_{ts} denote the joint density of $(\Delta v_t^{(c)}, \Delta v_s^{(c)})$ given realizations of $(\Delta v_t^{(d)}, \Delta v_s^{(d)})$.

Assumption 4 (Higher-order moments)**.** *For t, s in $2,\ldots,T$, $f_{ts}(\mathrm{v}_1^{(c)}, \mathrm{v}_2^{(c)}|\mathrm{v}_1^{(d)}, \mathrm{v}_2^{(d)})$ is bounded from above for all $(\mathrm{v}_1, \mathrm{v}_2)$ in $\mathcal{V}_t \otimes \mathcal{V}_s$ and is strictly positive for $(\mathrm{v}_1, \mathrm{v}_2)$ in a neighborhood of zero. For all such (t, s), $\mathscr{E}[z_t\omega_{\mathrm{v}_1}(\Delta v_t)\Delta\xi_t\Delta\xi_s\omega_{\mathrm{v}_2}(\Delta v_s)z_s']$ and $f_{ts}(\mathrm{v}_1^{(c)}, \mathrm{v}_2^{(c)}|\mathrm{v}_1^{(d)}, \mathrm{v}_2^{(d)})$ are k-times continuously differentiable in $(\mathrm{v}_1^{(c)}, \mathrm{v}_2^{(c)})$, and the absolute moments $\mathscr{E}[\omega_{\mathrm{v}}(\Delta v_t)\|z_t\Delta x_t'\|^3]$ and $\mathscr{E}[\omega_{\mathrm{v}}(\Delta v_t)\|z_t\Delta y_t\|^3]$ are finite. The matrix $\mathscr{E}[Z'\Omega_0\mathrm{D}\varepsilon\varepsilon'\mathrm{D}'\Omega_0 Z]$ is positive definite and nonsingular.*

This condition will ensure that, when properly scaled, $S_{Z\xi}$ converges to a zero-mean Gaussian process.

In line with previously introduced notation, let F_2 be an $\sum_{t=2}^{T}\ell(z_t) \times \sum_{t=2}^{T}\ell(z_t)$ matrix consisting of $(T-1)^2$ blocks, with the $(t-1, s-1)$th block given by $\iota_{\ell(z_t)}\iota_{\ell(z_s)}' f_{ts}(0|0)$. Define

$$\Upsilon \equiv \mathscr{E}\left[Z'\Omega_0\mathrm{D}\varepsilon\varepsilon'\mathrm{D}'\Omega_0 Z\right] \odot F_2 \int \mathrm{k}(\eta)^2\mathrm{d}\eta. \tag{3.6}$$

Then, under Assumptions 1–4, $\sqrt{n\sigma_n^{\ell}}S_{Z\xi} \xrightarrow{d} \mathcal{N}(0,\Upsilon)$. From this, the next theorem can be derived.

Theorem 2 (Normality). *Let Assumptions 1–4 hold. If G and Γ are positive-definite matrices so that $G \xrightarrow{p} \Gamma$ as $n \to \infty$ and Γ is non-stochastic, then*

$$\sqrt{n\sigma_n^\ell}\big(\theta_n - \theta_0\big) \xrightarrow{d} \mathcal{N}\big(0, \mathscr{V}(\Gamma)\big)$$

for $\mathscr{V}(\Gamma) \equiv \big[\Sigma_{ZX}'\Gamma\Sigma_{ZX}\big]^{-1}\big[\Sigma_{ZX}'\Gamma'\Upsilon\Gamma\Sigma_{ZX}\big]\big[\Sigma_{ZX}'\Gamma\Sigma_{ZX}\big]^{-1}$ *as* $n \to \infty$.

Proof. Given that $S_{ZX} \xrightarrow{p} \Sigma_{ZX}$ and $G \xrightarrow{p} \Gamma$ as $n \to \infty$, it remains to show that (i) $\sqrt{n\sigma_n^\ell}\mathscr{E}[S_{Z\xi}] = o(1)$; and (ii) $\sqrt{n\sigma_n^\ell}\big(S_{Z\xi} - \mathscr{E}[S_{Z\xi}]\big) \xrightarrow{d} \mathcal{N}(0, \Upsilon)$. Slutzky's theorem will take us the rest of the way.

Commence with (i). By a kth-order expansion of a typical block of $\mathscr{E}[S_{Z\xi}]$ around $\Delta v_t^{(c)} = 0$,

$$\bigg\|\frac{1}{\sigma_n^\ell}\mathscr{E}\Big[z_t\Delta\xi_t \mathrm{k}\Big(\frac{\Delta v_t^{(c)}}{\sigma_n}\Big)1(\Delta v_t^{(d)} = 0)\Big]\bigg\| = \Big\|\mathscr{E}\big[\omega_0(\Delta v_t)z_t\Delta\varepsilon_t\big]\ f_t(0|0)\Big\| + \mathcal{O}(\sigma_n^k).$$

The first right-hand side term above is zero by instrument validity. The order of magnitude of the remainder term follows by virtue of the higher-order kernel and the associated smoothness conditions. Consequently, $\sqrt{n\sigma_n^\ell}\mathscr{E}[S_{Z\xi}] = \sqrt{n\sigma_n^\ell}\mathcal{O}(\sigma_n^k) = o(1)$.

To verify (ii) we can follow Honoré and Kyriazidou (2000) in checking that the regularity conditions of Lyapunov's central limit theorem for double arrays hold. Write

$$\sqrt{n\sigma_n^\ell}c'\big(S_{Z\xi} - \mathscr{E}[S_{Z\xi}]\big) = \frac{1}{\sqrt{n\sigma_n^\ell}}\sum_{i=1}^n c'\big(q_i - \mathscr{E}[q]\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^N r_i$$

for any vector of constants c for which $c'c = 1$. Clearly, $\mathscr{E}[r] = 0$ while $\mathscr{E}[rr'] = \sigma_n^{-\ell}c'\mathscr{E}[qq']c - \sigma_n^{-\ell}c'\mathscr{E}[q]\mathscr{E}[q']c$. The two expectations in this variance are finite. Moreover, the second term is $\mathcal{O}(\sigma_n^\ell) = o(1)$ from above. For the first term, $\mathscr{E}[qq']$ has typical block

$$\frac{1}{\sigma_n^\ell}c'\mathscr{E}\Big[z_t\mathrm{k}\Big(\frac{\Delta v_t^{(c)}}{\sigma_n}\Big)1(\Delta v_t^{(d)} = 0)\Delta\xi_t\Delta\xi_s 1(\Delta v_s^{(d)} = 0)\mathrm{k}\Big(\frac{\Delta v_s^{(c)}}{\sigma_n}\Big)z_s'\Big]c.$$

Smoothness and dominated convergence again imply that, as $n \to \infty$, such blocks converge to

$$c'\mathscr{E}\big[z_t\omega_0(\Delta v_t)\Delta\varepsilon_t\Delta\varepsilon_s\omega_0(\Delta v_s)z_s\big]c\ f_{ts}(0|0)\int \mathrm{k}(\eta)^2\mathrm{d}\eta.$$

Thus, $\overline{\mathscr{E}}[rr'] = \sigma_n^\ell c'\overline{\mathscr{E}}[qq']c = c'\Upsilon c$. Lastly, it is easy to show that the boundedness of the kernel and of the absolute moments in Assumption 4 imply that, for any α in $(0,1)$,

$$\sum_{i=1}^n \mathscr{E}\Big[\Big\|\frac{r}{\sqrt{n}}\Big\|^{2+\alpha}\Big] = \mathcal{O}\Big(\frac{1}{\sqrt{n\sigma_n^\ell}}\Big)^\alpha = o(1).$$

Lyapunov's theorem may then be applied to obtain statement (ii) and to complete the proof. □

By standard optimality theory for GMM estimators (Hansen, 1982), the efficient weighting scheme for the moment conditions is obtained on setting G proportional to the inverse of a consistent estimate of Υ. Doing so leads to the following derivative-result to Theorem 2.

Corollary 1 (Optimally-weighted GMM)**.** *Let the conditions for Theorem 2 hold for estimators θ_n^* and θ_n^{**} that use a weight matrices G^* and G^{**} that satisfy $G^* \xrightarrow{p} \Gamma$ and $G^{**} \xrightarrow{p} \Upsilon^{-1}$ as $n \to \infty$, respectively Then,*

$$\sqrt{n\sigma_n^\ell}\big(\theta_n^{**} - \theta_0\big) \xrightarrow{d} \mathcal{N}\big(0, \mathscr{V}\big), \quad \mathscr{V} \equiv \mathscr{V}(\Upsilon^{-1}) = \big[\Sigma_{ZX}'\Upsilon^{-1}\Sigma_{ZX}\big]^{-1}$$

Furthermore, $\mathscr{V}(\Gamma) \geq \mathscr{V}$ in the matrix sense.

Proof. The expression for $\mathscr{V}$ follows from Theorem 2 and a small calculation. The efficiency result is readily verified on noting that

$$\mathscr{V}(\Gamma) - \mathscr{V} = B'[I - A(A'A)^{-1}A']B$$

for $A \equiv \Upsilon^{-1/2}\Sigma_{ZX}$ and $B \equiv [\Sigma_{ZX}'\Gamma\Sigma_{ZX}]^{-1}[\Sigma_{ZX}'\Gamma\Upsilon^{1/2}]$, and that $[I - A(A'A)^{-1}A']$ is an idempotent matrix. □

The last ingredient needed for inference procedures on the basis of θ_n to become operational is a consistent estimate of Υ in (3.6), say U. Then, for any properly formed G, and for $G = U^{-1}$ in particular, the consistency of a plug-in estimate for $\mathscr{V}(\Gamma)$ follows immediately. An estimate of $\mathscr{V}$, for example, is the inverse of $S_{ZX}'U^{-1}S_{ZX}$. In the special case of just-identification, such as for the weighted least-squares estimator obtained on setting $Z_i = \mathrm{D}X_i$, the matrix S_{ZX} is square and symmetric, and we obtain the familiar 'sandwich-form' estimator $S_{ZX}^{-1}US_{ZX}^{-1}$. It suffices to show that the general estimator of Υ defined as

$$\frac{\sigma_n^\ell}{n}\sum_{i=1}^n Z_i'W_i\mathrm{D}e_ie_i'\mathrm{D}'W_iZ_i \tag{3.7}$$

for given residuals e_i is consistent as $n \to \infty$. Given the efforts made so far, showing this result poses no additional difficulty.

Theorem 3. *Let Assumptions 1–4 hold, let θ_n^* be an initial estimator to which Theorem 1 applies, and define $\xi_i^* \equiv y_i - X_i\theta_n^*$. Then, $U \xrightarrow{p} \Upsilon$ as $n \to \infty$ for U as defined in (3.7) and formed with $e_i = \xi_i^*$.*

Proof. The proof uses the same arguments as those used to prove Theorem 2. □

Under conditional homoskedasticity of ε_i, and in the absence of serial correlation, setting $U = n^{-1}\sum_{i=1}^{n} Z_i' W_i \mathrm{DD}' W_i Z_i$ will give rise to the optimally-weighted GMM estimator. Alternatively, a two-step estimator would use the residuals, ξ_i^*, to form U as in (3.7). More generally, an iterated GMM procedure—as discussed in Hall (2005, pp. 90) and also elsewhere—repeats the two-step procedure until convergence. The estimate so formed is independent of the initial choice for G. Of course, the same comments about the effect of the variability of G on the small-sample performance of GMM estimators apply to the local GMM estimator introduced here. In light of this, it might be of interest to work out a correction term analogous to the one derived by Windmeijer (2005) for linear GMM estimators.

An alternative to an argument based on such an exercise in higher-order asymptotics is to consider a continuously-updated (CU) GMM routine as initially introduced by Hansen, Heaton, and Yaron (1996). Here, such an estimator takes the form

$$\arg\min_{\theta} S_{Z\xi}(\theta)' U(\theta)^{-1} S_{Z\xi}(\theta)$$

where $U(\theta)$ is U for a given θ. Such a modification makes the minimization problem nonlinear in θ but does not distort the limiting distribution of the point estimates, so that both Theorem 1 and Theorem 2 continue to apply. The CU estimator is a device originally suggested to mitigate the finite-sample bias of efficient GMM estimators that utilize overidentifying restrictions. Although it is not known whether this estimator has any finite moments, it is generally found to perform better in simulations in terms of bias than the corresponding two-step estimator. However, it also tends to exhibit a higher dispersion. Another useful feature of the CU estimator that explains its improved centering property is its interpretation as a jackknife estimator; see Donald and Newey (2000) for this. A final potentially benevolent effect of the CU approach is its invariance with respect to how the moment conditions are scaled. For an illustration and discussion of the CU estimator in the conventional panel data context, see Arellano (2003, pp. 73 and pp. 171–172).

Like in the standard setting, the efficiently-weighted local GMM minimand, when evaluated at θ_n^{**}, can be used to test overidentifying restrictions. The limiting result for the corresponding J-statistic is that

$$n\sigma_n^{\ell}\, S_{Z\xi}(\theta_n^{**})' U^{-1} S_{Z\xi}(\theta_n^{**}) \xrightarrow{d} \chi^2(m - \ell(x))$$

under Assumptions 1–4, where m denotes the number of columns of Z. The proof is immediate from Corollary 1. The result also extends straighforwardly to the CU estimator.

IV A MONTE CARLO STUDY

The performance of the local-GMM estimator was assessed through simulation exercises. This section reports results for static models estimated through kernel-weighted least squares and for dynamic models estimated through a local version of the first-differenced GMM procedure as originally introduced by Holtz-Eakin, Newey, and Rosen (1988) and Arellano and Bond (1991).

I use different design variations to be able to accentuate different findings in the static and the dynamic cases. The statistics used to evaluate the estimator's performance are the mean- and the median bias as measures of centrality, the standard deviation (STD) and the interquartile range (IQR) as indicators of dispersion, the root mean-squared error (RMSE) as a criterion for estimator risk, and coverage rates of two-sided 95% confidence intervals for inference evaluation. In the tables below CI(A) refers to coverage rates obtained through the use of plug-in estimates of the asymptotic variance. The entries for CI(B) are bootstrap-based acceptance frequencies using the percentile method over 39 bootstrap replications. All descriptive statistics for each design point were obtained over $10,000$ simulation runs, with all variables and incidental parameters redrawn in each iteration.

Below, θ_n^* refers to the one-step local-GMM estimator while θ_n^{**} denotes the two-step version. I experimented with iterating the two-step procedure but found no significant gain from doing so. I also ran simulation experiments with the CU estimator. While I found it to have reasonably small bias, its standard deviation was often very high, which is in line with results found in the cross-sectional literature. Here, I prefer to report results on jackknifed versions of the local-GMM estimator, which is an alternative route to constructing bias-corrected estimators. The jackknifed estimators take the usual delete-one form

$$\theta_n^\dagger \equiv n\theta_n^* - \frac{n-1}{n}\sum_{i=1}^n \theta_{n,-i}^* \quad \text{and} \quad \theta_n^{\dagger\dagger} \equiv n\theta_n^{**} - \frac{n-1}{n}\sum_{i=1}^n \theta_{n,-i}^{**}$$

where $\theta_{n,-i}^*$ and $\theta_{n,-i}^{**}$ are the one-step and two-step estimators obtained from estimating θ_0 from the subpanel obtained on deleting the ith time series from the full panel. In the conventional setting, it is known that the bias of the first-differenced GMM is $\mathcal{O}(n^{-1})$; see Arellano (2004). It is therefore reasonable to expect that, for k chosen sufficiently large, the bias of the local-GMM estimator has a leading term that is $\mathcal{O}(n^{-1})$. The jackknifed estimators $\theta_n^\dagger$ and $\theta_n^{\dagger\dagger}$ subtract a nonparametric estimate of the leading bias term from θ_n^* and θ_n^{**}, respectively.

Table 1: Monte Carlo results for model (4.8)–(4.9) under independence ($\rho_0 = 0$).

WEIGHT			MEAN BIAS		MEDIAN BIAS		STD		IQR		RMSE		CI(A)		CI(B)	
T	k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$
3	Normal	small	−.0006	−.0009	.0004	−.0009	.1609	.1627	.2093	.2114	.1609	.1627	.9099	.9058	.9193	.9129
3	Normal	medium	−.0025	−.0026	−.0025	−.0024	.1006	.1008	.1343	.1338	.1007	.1008	.9323	.9318	.9299	.9279
3	Normal	large	−.0038	−.0039	−.0043	−.0047	.0962	.0964	.1262	.1266	.0963	.0965	.9337	.9339	.9326	.9304
3	Epanech.	small	.0021	.0014	−.0002	.0000	.2435	.2507	.3106	.3178	.2435	.2507	.8851	.8705	.9184	.9137
3	Epanech.	medium	−.0018	−.0021	−.0022	−.0024	.1369	.1378	.1790	.1797	.1369	.1378	.9246	.9208	.9284	.9274
3	Epanech.	large	−.0029	−.0030	−.0030	−.0034	.0947	.0953	.1266	.1269	.0948	.0953	.9373	.9342	.9361	.9357
3	Quartic	small	.0017	.0006	−.0025	−.0013	.2671	.2777	.3392	.3515	.2671	.2777	.8754	.8557	.9161	.9073
3	Quartic	medium	−.0013	−.0015	−.0008	−.0006	.1485	.1494	.1954	.1955	.1485	.1494	.9185	.9149	.9254	.9206
3	Quartic	large	−.0026	−.0028	−.0024	−.0023	.0969	.0971	.1279	.1280	.0969	.0971	.9353	.9345	.9340	.9337
3	Cosine	small	.0020	.0013	−.0007	−.0012	.2468	.2544	.3151	.3246	.2468	.2544	.8831	.8675	.9194	.9121
3	Cosine	medium	−.0017	−.0020	−.0019	−.0021	.1386	.1394	.1812	.1817	.1386	.1394	.9245	.9212	.9282	.9263
3	Cosine	large	−.0029	−.0030	−.0025	−.0025	.0947	.0952	.1265	.1266	.0948	.0952	.9366	.9345	.9362	.9354
6	Normal	small	−.0020	−.0019	−.0021	−.0019	.1050	.1054	.1396	.1404	.1050	.1054	.9377	.9364	.9342	.9320
6	Normal	medium	−.0013	−.0012	.0000	−.0001	.0670	.0671	.0880	.0885	.0670	.0671	.9414	.9412	.9372	.9365
6	Normal	large	−.0011	−.0012	−.0005	−.0006	.0704	.0706	.0950	.0954	.0705	.0706	.9439	.9430	.9390	.9389
6	Epanech.	small	−.0027	−.0025	−.0024	−.0025	.1538	.1552	.2028	.2028	.1538	.1552	.9263	.9222	.9329	.9285
6	Epanech.	medium	−.0017	−.0017	−.0005	−.0003	.0901	.0905	.1185	.1189	.0901	.0905	.9391	.9375	.9344	.9333
6	Epanech.	large	−.0012	−.0013	.0004	.0002	.0646	.0651	.0861	.0865	.0646	.0651	.9429	.9411	.9385	.9400
6	Quartic	small	−.0030	−.0028	−.0021	−.0020	.1672	.1689	.2221	.2241	.1672	.1689	.9215	.9171	.9309	.9245
6	Quartic	medium	−.0018	−.0018	−.0015	−.0015	.0973	.0975	.1288	.1290	.0973	.0975	.9391	.9377	.9335	.9321
6	Quartic	large	−.0012	−.0012	.0004	.0001	.0650	.0652	.0858	.0859	.0650	.0652	.9416	.9417	.9377	.9365
6	Cosine	small	−.0028	−.0025	−.0024	−.0025	.1557	.1571	.2053	.2061	.1557	.1571	.9256	.9212	.9328	.9280
6	Cosine	medium	−.0018	−.0017	−.0006	−.0007	.0912	.0915	.1199	.1206	.0912	.0915	.9390	.9370	.9342	.9326
6	Cosine	large	−.0012	−.0013	.0003	.0002	.0644	.0648	.0857	.0864	.0644	.0648	.9420	.9409	.9380	.9382
9	Normal	small	.0008	.0008	.0015	.0014	.0837	.0838	.1103	.1107	.0837	.0838	.9382	.9375	.9325	.9321
9	Normal	medium	.0002	.0002	.0001	.0003	.0543	.0544	.0736	.0736	.0543	.0544	.9441	.9441	.9393	.9389
9	Normal	large	−.0008	−.0008	−.0018	−.0019	.0623	.0624	.0831	.0832	.0623	.0624	.9391	.9388	.9361	.9355
9	Epanech.	small	.0004	.0002	.0013	.0011	.1223	.1230	.1596	.1604	.1223	.1229	.9317	.9305	.9329	.9329
9	Epanech.	medium	.0009	.0010	.0010	.0012	.0717	.0719	.0948	.0960	.0717	.0719	.9431	.9429	.9393	.9396
9	Epanech.	large	−.0002	−.0002	−.0006	−.0004	.0541	.0544	.0720	.0721	.0541	.0544	.9442	.9410	.9395	.9395
9	Quartic	small	.0003	.0002	.0008	.0012	.1334	.1340	.1746	.1755	.1334	.1340	.9273	.9245	.9298	.9262
9	Quartic	medium	.0009	.0009	.0014	.0016	.0775	.0775	.1025	.1028	.0775	.0775	.9403	.9402	.9338	.9337
9	Quartic	large	.0000	.0000	.0004	.0002	.0534	.0536	.0708	.0712	.0534	.0536	.9443	.9432	.9402	.9404
9	Cosine	small	.0004	.0002	.0014	.0008	.1240	.1246	.1614	.1626	.1240	.1245	.9311	.9292	.9325	.9315
9	Cosine	medium	.0009	.0010	.0014	.0015	.0726	.0727	.0961	.0967	.0726	.0727	.9431	.9429	.9378	.9384
9	Cosine	large	−.0002	−.0002	−.0005	−.0005	.0537	.0540	.0713	.0713	.0537	.0540	.9442	.9416	.9387	.9394

$n = 100$, $\theta_0 = .5$, 10,000 Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

Static models. The first model considered has scalar x and v influencing the outcome variable through

$$y_{it} = x_{it}\theta_0 + \vartheta_i(v_{it}) + \varepsilon_{it}, \quad \vartheta_i(v_{it}) = \gamma_i^{(1)} + \gamma_i^{(2)} v_{it} - \gamma_i^{(3)} v_{it}^2. \tag{4.8}$$

Here, $\gamma_i^{(1)} \sim \mathcal{N}(0,1)$ are traditional fixed effects while $\gamma_i^{(2)} \sim \mathcal{N}(0,2)$ and $\gamma_i^{(3)} \sim \mathcal{N}(0,.75)$ appear in an interactive fashion. The form of ϑ_i allows for both concave and convex response functions. The dynamics of the covariates x and v were set to be

$$x_{it} = -.3\gamma_i^{(1)} + .5x_{it-1} + \beta_i^{(1)} v_{it} + \mathcal{N}(0,1) \quad \text{and} \quad v_{it} = .3\gamma_i^{(1)} + \beta_i^{(2)} v_{it-1} + \zeta_{it}, \tag{4.9}$$

respectively. In (4.9), $\beta_i^{(1)} \sim \mathcal{N}(0,1)$ and $\beta_i^{(2)} \sim \mathcal{U}[.20,.99]$, so that both strongly- and weakly-persistent data sequences occur. Because differencing is executed over time the within-group variation in ϑ_i is much more important for the performance of the weighting than is the between-group variation. The start-up values for the covariate time series were generated as $x_{i0} \sim -.3\gamma_i^{(1)} + \mathcal{N}(0,1)$ and $v_{i0} \sim .3\gamma_i^{(1)} + \mathcal{N}(0,1)$. The disturbances are bivariate-normal variates, each with unit variance and a correlation coefficient ρ_0 which was 0 in the first design and .4 in the second. The latter case leads to x, v, and ε being contemporaneously correlated, making x endogenous in the cross-sectional sense. Observe also that x and v are dependent, so that a naive fixed-effect estimator will be inconsistent for θ_0.

Tables 1 and 2 contain the results for the kernel-weighted first-differenced least-squares estimator for $\rho_0 = 0$ and $\rho_0 = .4$, respectively. Because, here, there are no overidentifying restrictions, both θ_n^* and θ_n^{**}, and $\theta_n^\dagger$ and $\theta_n^{\dagger\dagger}$ coincide. Throughout, θ_0 was fixed at .5.[4] The dimensions of the panels generated were $n = 100$ and $T \in \{3,6,9\}$. Results are reported for the standard-normal kernel (Normal), the second-order Epanechnikov kernel (Epanechnikov), the quartic kernel (Quartic), and the cosine kernel (Cosine). Three different bandwidth choices are reported on; $5n^{-3/4}$ (small), $15n^{-3/4}$ (medium), and $45n^{-3/4}$ (large).

Overal, Table 1 reports good performance of θ_n^*. Both the mean- and the median bias are virtually zero throughout. The jackknifed estimator has a bias of the same magnitude. Not surprisingly, $\theta_n^\dagger$ is more volatile. The difference in the STD and IQR is, however, very small. As a consequence, the RMSE of both estimators is identical to their STD up to the first few decimal digits. Both CI(A) and CI(B) report solid coverage rates, with neither dominating the other.[5] The actual size is generally somewhat larger

[4]Experiments with different values for θ_0 gave qualitatively similar results.

[5]Here, CI(A) was obtained by means of the heteroskedasticity-robust variance-covariance estimator, as displayed in the text.

than the nominal size of .05, which is in line with the usual behavior found in Monte Carlo work elsewhere. The impact of the choice of the kernel and the bandwidth is strongest on the measures of spread. But no design constellation uniformly dominates another.

When turning to the design with contemporaneous dependence between x and ε in Table 2 the overal picture changes little. Both estimators continue to have an empirical distribution that is roughly correctly centered; their variance changes little, too. Thus, confidence intervals still provide approximately correct coverage rates. No consistent pattern emerges when comparing the numbers across the tables.

Dynamic models. This subsection deals with the following variation on (4.8)–(4.9).

$$y_{it} = y_{it-1}\theta_0 + \vartheta_i(v_{it}) + \varepsilon_{it}, \quad \vartheta_i(v_{it}) = \gamma_i^{(1)} + \gamma_i^{(2)} v_{it} - \gamma_i^{(3)} v_{it}^2, \tag{4.10}$$

with

$$v_{it} = .3\gamma_i^{(1)} + \beta_i^{(2)}\big(\rho_0 v_{it-1} + (1-\rho_0)\,\mathcal{U}[-1,1]\big) + \zeta_{it} \tag{4.11}$$

for $\rho_0 \in \{0,1\}$ and i.i.d. errors $(\varepsilon_{it}, \zeta_{it}) \sim \mathcal{N}(0, I)$. The processes were generated with a burn in of 500 time periods, using

$$v_{i(-501)} \sim .3\gamma_i^{(1)} + \mathcal{N}(0,1) \quad \text{and} \quad y_{i(-501)} \sim -.3\gamma_i^{(1)} + \vartheta_i(v_{i(-501)}) + \mathcal{N}(0,1)$$

as startup values. The local-GMM estimator of θ_0 à la Arellano and Bond (1991) uses all lagged levels of y_{it-1} as instruments for Δy_{it-1} in the equations in first differences.

Like with the conventional estimator, the small-sample bias of θ_n^* and θ_n^{**} will be a function of the number of moments used as well as the instrument strength. The former relates directly to the size of the panel (see, e.g., Alvarez and Arellano, 2003), the latter is partly driven by the closeness of the autoregressive parameter θ_0 to unity, with the instruments becoming irrelevant when $\theta_0 = 1$. In addition, coverage intervals based on the two-step estimator that are set up by means of analytical standard errors would be expected to be too small, leading to nominal rejection frequencies that are too high compared to a chosen significance level. Table 7 below contains the average of the estimated standard errors of both θ_n^* and θ_n^{**} by means of the analytical formulae and the bootstrap, together with the empirical standard deviation for a selection of the Monte Carlo experiments that follow. It can be observed that, for both estimators, the plug-in estimates are too low on average. The bootstrap-based standard errors are much more in line with the variability of the estimators as actually observed over the Monte Carlo runs. For this reason, I focus on the coverage rates of the confidence intervals based on resampling below.

Table 2: Monte Carlo results for model (4.8)–(4.9) under dependence ($\rho_0 = .4$).

WEIGHT			MEAN BIAS		MEDIAN BIAS		STD		IQR		RMSE		CI(A)		CI(B)	
T	k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$	θ_n^*	$\theta_n^\dagger$
3	Normal	small	−.0003	−.0006	−.0010	−.0009	.1532	.1554	.2046	.2079	.1532	.1554	.9184	.9107	.9243	.9192
3	Normal	medium	−.0011	−.0012	−.0002	−.0004	.1000	.1003	.1338	.1339	.1000	.1003	.9350	.9330	.9296	.9294
3	Normal	large	−.0036	−.0038	−.0039	−.0040	.1060	.1063	.1400	.1400	.1061	.1064	.9351	.9350	.9327	.9324
3	Epanech.	small	.0018	.0012	.0000	.0006	.2302	.2387	.2919	.2967	.2302	.2387	.8897	.8757	.9213	.9162
3	Epanech.	medium	−.0002	−.0004	−.0004	.0000	.1328	.1341	.1769	.1795	.1328	.1341	.9294	.9256	.9320	.9290
3	Epanech.	large	−.0023	−.0025	−.0026	−.0028	.0984	.0993	.1308	.1310	.0985	.0994	.9312	.9293	.9319	.9321
3	Quartic	small	.0015	.0007	−.0007	.0001	.2532	.2658	.3188	.3319	.2532	.2657	.8784	.8603	.9206	.9103
3	Quartic	medium	−.0004	−.0007	−.0012	−.0015	.1428	.1441	.1893	.1919	.1428	.1440	.9256	.9195	.9273	.9241
3	Quartic	large	−.0016	−.0018	−.0008	−.0008	.0982	.0985	.1313	.1316	.0982	.0985	.9338	.9321	.9311	.9299
3	Cosine	small	.0017	.0012	.0002	.0006	.2335	.2423	.2954	.3014	.2334	.2423	.8889	.8728	.9206	.9133
3	Cosine	medium	−.0002	−.0004	−.0003	−.0007	.1342	.1354	.1786	.1825	.1342	.1354	.9290	.9241	.9305	.9280
3	Cosine	large	−.0022	−.0023	−.0025	−.0029	.0980	.0987	.1309	.1309	.0980	.0987	.9307	.9298	.9317	.9315
6	Normal	small	.0010	.0009	.0010	.0010	.1010	.1014	.1340	.1347	.1010	.1014	.9308	.9293	.9302	.9283
6	Normal	medium	−.0007	−.0007	−.0004	−.0005	.0664	.0665	.0899	.0899	.0664	.0665	.9406	.9401	.9365	.9351
6	Normal	large	−.0014	−.0015	−.0008	−.0011	.0799	.0800	.1100	.1102	.0799	.0801	.9427	.9423	.9395	.9390
6	Epanech.	small	.0007	.0005	.0033	.0023	.1475	.1489	.2000	.1997	.1475	.1489	.9198	.9153	.9258	.9237
6	Epanech.	medium	.0007	.0007	−.0001	.0005	.0871	.0877	.1159	.1169	.0871	.0877	.9333	.9321	.9316	.9318
6	Epanech.	large	−.0012	−.0012	−.0007	−.0007	.0675	.0682	.0900	.0910	.0675	.0682	.9413	.9381	.9370	.9354
6	Quartic	small	.0001	−.0003	.0018	.0023	.1614	.1633	.2136	.2160	.1614	.1633	.9132	.9083	.9224	.9167
6	Quartic	medium	.0009	.0009	.0006	.0006	.0938	.0941	.1249	.1252	.0938	.0941	.9318	.9311	.9311	.9297
6	Quartic	large	−.0010	−.0010	−.0005	−.0006	.0660	.0662	.0883	.0885	.0660	.0662	.9403	.9385	.9384	.9377
6	Cosine	small	.0006	.0003	.0023	.0018	.1496	.1510	.2017	.2026	.1496	.1510	.9185	.9145	.9247	.9215
6	Cosine	medium	.0007	.0007	.0002	.0000	.0881	.0885	.1171	.1177	.0881	.0885	.9336	.9326	.9318	.9309
6	Cosine	large	−.0011	−.0011	−.0009	−.0007	.0669	.0675	.0893	.0906	.0669	.0675	.9419	.9390	.9367	.9363
9	Normal	small	.0012	.0012	.0010	.0012	.0803	.0804	.1074	.1083	.0803	.0804	.9356	.9351	.9305	.9305
9	Normal	medium	.0010	.0010	.0017	.0017	.0541	.0541	.0724	.0725	.0541	.0541	.9411	.9412	.9363	.9361
9	Normal	large	−.0005	−.0006	−.0007	−.0006	.0715	.0716	.0955	.0956	.0715	.0716	.9394	.9393	.9357	.9362
9	Epanech.	small	.0009	.0009	.0006	.0000	.1158	.1166	.1547	.1560	.1158	.1166	.9302	.9290	.9325	.9309
9	Epanech.	medium	.0013	.0013	.0018	.0014	.0693	.0696	.0936	.0938	.0693	.0696	.9393	.9388	.9345	.9362
9	Epanech.	large	.0006	.0005	.0007	.0004	.0568	.0572	.0764	.0769	.0568	.0572	.9443	.9431	.9397	.9399
9	Quartic	small	.0005	.0004	−.0001	−.0001	.1259	.1268	.1679	.1690	.1259	.1268	.9264	.9245	.9313	.9271
9	Quartic	medium	.0012	.0012	.0015	.0014	.0746	.0747	.1001	.1005	.0746	.0747	.9371	.9366	.9333	.9336
9	Quartic	large	.0008	.0008	.0010	.0010	.0544	.0546	.0731	.0730	.0544	.0546	.9419	.9416	.9380	.9378
9	Cosine	small	.0008	.0008	.0001	−.0001	.1173	.1181	.1573	.1579	.1173	.1180	.9304	.9279	.9320	.9300
9	Cosine	medium	.0013	.0013	.0014	.0014	.0701	.0703	.0947	.0952	.0701	.0703	.9389	.9386	.9350	.9357
9	Cosine	large	.0006	.0006	.0006	.0004	.0561	.0564	.0755	.0763	.0561	.0564	.9453	.9425	.9397	.9390

$n = 100$, $\theta_0 = .5$, 10,000 Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

Table 3: Monte Carlo results for model (4.10)–(4.11); $\rho_0 = 0$, $T = 3$.

WEIGHT		MEAN BIAS				MEDIAN BIAS			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	$-.0269$	.0138	$-.0244$	.0171	$-.0418$	$-.0109$	$-.0377$	$-.0084$
Normal	medium	$-.0395$	.0127	$-.0292$	.0270	$-.0531$	$-.0131$	$-.0456$	$-.0087$
Normal	large	$-.0486$	.0116	$-.0351$	.0244	$-.0648$	$-.0167$	$-.0546$	$-.0128$
Epanech.	small	$-.0454$	.0207	$-.0449$	.0143	$-.0646$	$-.0306$	$-.0649$	$-.0317$
Epanech.	medium	$-.0298$	.0158	$-.0238$	.0239	$-.0423$	$-.0080$	$-.0347$	$-.0036$
Epanech.	large	$-.0469$	.0119	$-.0341$	.0233	$-.0628$	$-.0155$	$-.0529$	$-.0130$
Quartic	small	$-.0540$	.0130	$-.0535$	.0138	$-.0736$	$-.0382$	$-.0759$	$-.0377$
Quartic	medium	$-.0268$	.0144	$-.0231$	.0174	$-.0402$	$-.0071$	$-.0327$	$-.0049$
Quartic	large	$-.0442$	.0123	$-.0324$	.0208	$-.0603$	$-.0140$	$-.0486$	$-.0094$
WEIGHT		STD				IQR			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.2430	.3425	.2472	.4094	.2786	.2824	.2924	.3199
Normal	medium	.2702	.3367	.2727	.5898	.3137	.3421	.3179	.3526
Normal	large	.2915	.3629	.2928	.4589	.3390	.3646	.3379	.3876
Epanech.	small	.2998	.5632	.3069	.5004	.3329	.3728	.3535	.4158
Epanech.	medium	.2519	.3360	.2540	.4593	.2913	.3095	.2931	.3231
Epanech.	large	.2878	.3580	.2897	.4567	.3356	.3604	.3344	.3788
Quartic	small	.3274	.8390	.3382	.7210	.3550	.4042	.3729	.4498
Quartic	medium	.2435	.3068	.2462	.3443	.2816	.2984	.2896	.3181
Quartic	large	.2815	.3504	.2847	.4589	.3274	.3546	.3288	.3715
WEIGHT		RMSE				CI(B)			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.2445	.3428	.2484	.4098	.9332	.9354	.9506	.9524
Normal	medium	.2730	.3369	.2742	.5904	.9250	.9344	.9470	.9564
Normal	large	.2955	.3631	.2949	.4595	.9244	.9304	.9446	.9516
Epanech.	small	.3032	.5635	.3101	.5006	.9326	.9328	.9510	.9540
Epanech.	medium	.2536	.3364	.2551	.4599	.9320	.9366	.9496	.9554
Epanech.	large	.2916	.3582	.2917	.4572	.9250	.9308	.9452	.9516
Quartic	small	.3318	.8391	.3424	.7211	.9280	.9298	.9440	.9522
Quartic	medium	.2449	.3071	.2473	.3447	.9352	.9390	.9498	.9554
Quartic	large	.2849	.3506	.2865	.4593	.9260	.9318	.9468	.9524

$n = 100$, $\theta_0 = .5$, $10,000$ Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

An important difference between the least-squares estimator as considered in the previous subsection and the GMM estimator for the dynamic model is that it uses levels as instruments rather than differences. While this is also the case when ϑ_i does not depend on v, there might be additional finite-sample effects when v is autocorrelated. To see this, recall that local differencing removes $\Delta\vartheta_i(v_{it})$ from the moment conditions only asymptotically. For any finite n, therefore, lagged dependent variables will correlate with $w_{it}\Delta\xi_{it}$. To assess the importance of this serial correlation, I manipulate ρ_0 from zero to unity. When $\rho_0 = 0$, v is independent over time while, when $\rho_0 = 1$, it is first-order autocorrelated with some sequences being nearly integrated of order one.

The kernel and bandwidth choices from before were maintained, and n was equally kept at 100. To save on space, I report results only for a subset of the design variations. Additional results are available as supplementary material.

Table 4: Monte Carlo results for model (4.10)–(4.11); $\rho_0 = 1$, $T = 3$.

WEIGHT		MEAN BIAS				MEDIAN BIAS			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.0158	.0769	.0104	.0638	.0012	.0406	.0019	.0359
Normal	medium	.0563	.1500	.0519	.1634	.0382	.0991	.0385	.0925
Normal	large	.0604	.1774	.0545	.2262	.0430	.1069	.0372	.1015
Epanech.	small	−.0429	.0567	−.0474	.0191	−.0660	−.0211	−.0622	−.0208
Epanech.	medium	.0409	.1152	.0341	.0887	.0262	.0737	.0297	.0771
Epanech.	large	.0601	.1711	.0550	.2466	.0419	.1059	.0376	.0989
Quartic	small	−.0526	.0711	−.0547	.0570	−.0748	−.0312	−.0721	−.0330
Quartic	medium	.0287	.0994	.0223	.0772	.0149	.0564	.0144	.0584
Quartic	large	.0592	.1632	.0548	.2105	.0409	.1026	.0364	.0972
WEIGHT		STD				IQR			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.3178	.7237	.3294	1.0133	.3191	.3472	.3254	.3806
Normal	medium	.3921	1.2205	.4015	1.7770	.3674	.4143	.3614	.4494
Normal	large	.4331	1.0987	.4475	2.5873	.4008	.4526	.3903	.4816
Epanech.	small	.3578	1.3454	.3903	2.0493	.3571	.4037	.3686	.4507
Epanech.	medium	.3396	.7551	.3578	1.2447	.3449	.3779	.3309	.4035
Epanech.	large	.4251	1.0993	.4340	3.7695	.3960	.4467	.3855	.4750
Quartic	small	.3814	2.5693	.4034	2.5677	.3720	.4307	.3900	.4859
Quartic	medium	.3190	.6842	.3356	1.0863	.3305	.3622	.3234	.3843
Quartic	large	.4127	1.1089	.4221	2.5327	.3884	.4375	.3791	.4613
WEIGHT		RMSE				CI(B)			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.3182	.7277	.3295	1.0152	.9512	.9648	.9516	.9654
Normal	medium	.3961	1.2296	.4049	1.7844	.9482	.9580	.9522	.9688
Normal	large	.4373	1.1128	.4508	2.5970	.9452	.9588	.9572	.9694
Epanech.	small	.3603	1.3465	.3931	2.0493	.9326	.9560	.9360	.9584
Epanech.	medium	.3421	.7638	.3594	1.2478	.9494	.9616	.9506	.9660
Epanech.	large	.4293	1.1125	.4374	3.7774	.9462	.9588	.9566	.9696
Quartic	small	.3850	2.5701	.4070	2.5682	.9310	.9514	.9358	.9604
Quartic	medium	.3203	.6914	.3363	1.0890	.9490	.9616	.9504	.9664
Quartic	large	.4169	1.1208	.4256	2.5413	.9460	.9580	.9544	.9692

$n = 100$, $\theta_0 = .5$, 10,000 Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

Table 3 gives the usual descriptive statistics for the case where $T = 3$, $\theta_0 = .5$, and $\rho_0 = 0$. Both the one- and two-step estimator have a reasonably small bias. The choice of k and σ_n seems particularly important here. The jackknifed estimators give an improvement in terms of centrality, but at a cost of a higher STD and RMSE. When n

is larger, however, this discrepancy vanishes. The IQR of $\theta_n^{\dagger}$ and $\theta_n^{\dagger\dagger}$ are, however, not much larger than those of θ_n^* and θ_n^{**} and also the coverage rates are comparable.

When ρ_0 is set to unity, the estimators' performance change. All estimators now have both a larger bias and a higher variability. Although the bias never skyrockets, the jackknife estimators do not outperform θ_n^* and θ_n^{**} in terms of any of the statistics reported, although they start doing so for larger n. Moreover, they are more biased than the original estimators, and their standard deviation is large. Nevertheless, the IQR remains relatively small and the coverage rates are close to .95.

Table 5: Monte Carlo results for model (4.10)–(4.11); $\rho_0 = 0$, $T = 6$.

WEIGHT		MEAN BIAS				MEDIAN BIAS			
k	σ_n	θ_n^*	$\theta_n^{\dagger}$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^{\dagger}$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	−.0558	−.0036	−.0560	−.0031	−.0548	−.0052	−.0541	−.0066
Normal	medium	−.0469	−.0041	−.0436	−.0033	−.0461	−.0033	−.0426	−.0034
Normal	large	−.0520	−.0062	−.0466	−.0040	−.0512	−.0070	−.0454	−.0040
Epanech.	small	−.1025	−.0136	−.1035	−.0139	−.1030	−.0202	−.1039	−.0235
Epanech.	medium	−.0469	−.0029	−.0460	−.0028	−.0473	−.0040	−.0462	−.0056
Epanech.	large	−.0509	−.0058	−.0459	−.0039	−.0505	−.0062	−.0444	−.0046
Quartic	small	−.1192	−.0184	−.1187	−.0165	−.1210	−.0279	−.1203	−.0291
Quartic	medium	−.0496	−.0028	−.0494	−.0025	−.0505	−.0053	−.0483	−.0035
Quartic	large	−.0493	−.0051	−.0449	−.0037	−.0485	−.0050	−.0439	−.0045
WEIGHT		STD				IQR			
k	σ_n	θ_n^*	$\theta_n^{\dagger}$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^{\dagger}$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.0999	.1168	.1092	.1395	.1323	.1541	.1435	.1812
Normal	medium	.0953	.1059	.1021	.1267	.1265	.1400	.1377	.1620
Normal	large	.1008	.1126	.1060	.1329	.1333	.1479	.1409	.1748
Epanech.	small	.1261	.1646	.1398	.1989	.1650	.2033	.1801	.2508
Epanech.	medium	.0955	.1085	.1033	.1304	.1254	.1399	.1364	.1690
Epanech.	large	.0997	.1111	.1052	.1315	.1321	.1467	.1406	.1722
Quartic	small	.1356	.1827	.1502	.2232	.1750	.2276	.1963	.2868
Quartic	medium	.0968	.1111	.1052	.1329	.1282	.1473	.1398	.1729
Quartic	large	.0979	.1089	.1040	.1295	.1310	.1450	.1389	.1687
WEIGHT		RMSE				CI(B)			
k	σ_n	θ_n^*	$\theta_n^{\dagger}$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^{\dagger}$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.1144	.1169	.1228	.1395	.8790	.9322	.8880	.9270
Normal	medium	.1062	.1059	.1110	.1268	.8990	.9322	.9126	.9308
Normal	large	.1134	.1128	.1157	.1329	.8962	.9280	.9102	.9290
Epanech.	small	.1625	.1652	.1739	.1994	.8272	.9178	.8450	.9216
Epanech.	medium	.1064	.1085	.1130	.1304	.8946	.9308	.9058	.9248
Epanech.	large	.1119	.1113	.1148	.1315	.8978	.9290	.9104	.9276
Quartic	small	.1805	.1836	.1914	.2238	.8036	.9126	.8246	.9168
Quartic	medium	.1087	.1112	.1162	.1329	.8856	.9318	.8974	.9244
Quartic	large	.1096	.1090	.1132	.1295	.8968	.9316	.9114	.9290

$n = 100$, $\theta_0 = .5$, 10,000 Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

Tables 5 and 6 show what happens when T is increased to 6. Of course, the STD and IQR shrink, and considerably so. For $\rho_0 = 0$, the bias of all estimators increases while, for $\rho_0 = 1$, the opposite happens. The coverage rates worsen slightly across both design variations. This effect is less pronounced for $\theta_n^\dagger$ and $\theta_n^{\dagger\dagger}$ than it is for θ_n^* and θ_n^{**}.

Table 6: Monte Carlo results for model (4.10)–(4.11); $\rho_0 = 1$, $T = 6$.

WEIGHT		MEAN BIAS				MEDIAN BIAS			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	−.0138	.0474	−.0212	.0376	−.0167	.0398	−.0197	.0331
Normal	medium	.0454	.1075	.0324	.0860	.0451	.1025	.0327	.0825
Normal	large	.0516	.1218	.0376	.0965	.0523	.1158	.0375	.0902
Epanech.	small	−.0897	−.0022	−.0933	−.0058	−.0922	−.0107	−.0930	−.0144
Epanech.	medium	.0230	.0810	.0113	.0641	.0206	.0738	.0114	.0603
Epanech.	large	.0510	.1197	.0372	.0949	.0514	.1137	.0373	.0890
Quartic	small	−.1078	−.0137	−.1099	−.0148	−.1093	−.0221	−.1115	−.0276
Quartic	medium	.0055	.0638	−.0043	.0502	.0032	.0568	−.0053	.0460
Quartic	large	.0496	.1158	.0360	.0921	.0497	.1102	.0368	.0871
WEIGHT		STD				IQR			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.1161	.1484	.1211	.1719	.1510	.1821	.1547	.2072
Normal	medium	.1195	.1488	.1160	.1624	.1548	.1835	.1536	.2019
Normal	large	.1269	.1603	.1215	.1724	.1654	.1967	.1605	.2167
Epanech.	small	.1357	.1906	.1487	.2325	.1788	.2328	.1948	.2800
Epanech.	medium	.1180	.1485	.1170	.1652	.1527	.1832	.1531	.2016
Epanech.	large	.1255	.1580	.1204	.1704	.1630	.1941	.1591	.2116
Quartic	small	.1416	.2037	.1562	.2495	.1842	.2458	.2029	.3024
Quartic	medium	.1167	.1475	.1184	.1670	.1524	.1833	.1531	.2014
Quartic	large	.1233	.1545	.1187	.1672	.1609	.1903	.1562	.2094
WEIGHT		RMSE				CI(B)			
k	σ_n	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$	θ_n^*	$\theta_n^\dagger$	θ_n^{**}	$\theta_n^{\dagger\dagger}$
Normal	small	.1169	.1558	.1230	.1760	.9190	.9122	.9250	.9202
Normal	medium	.1279	.1836	.1205	.1838	.8982	.8384	.9252	.8858
Normal	large	.1370	.2013	.1272	.1976	.8924	.8194	.9214	.8822
Epanech.	small	.1627	.1906	.1755	.2325	.8474	.9140	.8616	.9258
Epanech.	medium	.1202	.1691	.1175	.1772	.9202	.8818	.9320	.9102
Epanech.	large	.1355	.1982	.1260	.1950	.8930	.8206	.9224	.8834
Quartic	small	.1780	.2041	.1909	.2499	.8228	.9080	.8384	.9214
Quartic	medium	.1168	.1607	.1185	.1744	.9212	.9002	.9318	.9176
Quartic	large	.1329	.1931	.1240	.1908	.8936	.8234	.9240	.8848

$n = 100$, $\theta_0 = .5$, 10,000 Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

The last table in this subsection makes the comparison between the average estimated standard errors of θ_n^* and θ_n^{**} and their empirical standard deviation.

Table 7: A comparison between the estimated standard errors and the empirical standard deviation

			$\rho_0 = 0$						$\rho_0 = 1$					
DESIGN			SE (analytic)		SE (bootstrap)		STD		SE (analytic)		SE (bootstrap)		STD	
T	k	σ_n	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}
3	Normal	small	.1798	.2185	.2493	.2632	.2430	.2472	.2086	.2632	.3067	.3301	.3178	.3294
3	Normal	medium	.1670	.2359	.2698	.2831	.2702	.2727	.2020	.3081	.3598	.3840	.3921	.4015
3	Normal	large	.1703	.2513	.2888	.3021	.2915	.2928	.2094	.3353	.3849	.4094	.4331	.4475
3	Epanech.	small	.2384	.2631	.3116	.3345	.2998	.3069	.2731	.3012	.3463	.3728	.3578	.3903
3	Epanech.	medium	.1672	.2231	.2546	.2669	.2519	.2540	.1969	.2822	.3317	.3554	.3396	.3578
3	Epanech.	large	.1696	.2486	.2853	.2984	.2878	.2897	.2077	.3301	.3808	.4047	.4251	.4340
3	Quartic	small	.2622	.2810	.3325	.3564	.3274	.3382	.3000	.3183	.3606	.3902	.3814	.4034
3	Quartic	medium	.1711	.2185	.2498	.2633	.2435	.2462	.1993	.2684	.3162	.3393	.3190	.3356
3	Quartic	large	.1685	.2440	.2796	.2930	.2815	.2847	.2053	.3218	.3734	.3965	.4127	.4221
T	k	σ_n	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}
6	Normal	small	.0962	.0805	.0964	.1060	.0999	.1092	.1026	.0826	.1068	.1154	.1161	.1211
6	Normal	medium	.0783	.0741	.0922	.0987	.0953	.1021	.0827	.0779	.1083	.1119	.1195	.1160
6	Normal	large	.0781	.0765	.0963	.1023	.1008	.1060	.0826	.0810	.1141	.1169	.1269	.1215
6	Epanech.	small	.1330	.0999	.1224	.1367	.1261	.1398	.1407	.0985	.1265	.1410	.1357	.1487
6	Epanech.	medium	.0826	.0753	.0924	.0998	.0955	.1033	.0876	.0793	.1071	.1121	.1180	.1170
6	Epanech.	large	.0780	.0760	.0956	.1016	.0997	.1052	.0825	.0804	.1130	.1160	.1255	.1204
6	Quartic	small	.1445	.1043	.1294	.1444	.1356	.1502	.1521	.1018	.1316	.1467	.1416	.1562
6	Quartic	medium	.0880	.0773	.0935	.1019	.0968	.1052	.0936	.0805	.1063	.1129	.1167	.1184
6	Quartic	large	.0779	.0752	.0943	.1005	.0979	.1040	.0824	.0794	.1113	.1144	.1233	.1187
T	k	σ_n	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}	θ_n^*	θ_n^{**}
9	Normal	small	.0818	.0516	.0757	.0808	.0793	.0891	.0881	.0524	.0837	.0881	.0911	.0971
9	Normal	medium	.0640	.0452	.0696	.0732	.0735	.0804	.0681	.0461	.0794	.0819	.0921	.0922
9	Normal	large	.0633	.0459	.0719	.0752	.0763	.0823	.0672	.0469	.0826	.0848	.0962	.0956
9	Epanech.	small	.1125	.0622	.0955	.1020	.0995	.1115	.1200	.0610	.0989	.1052	.1045	.1166
9	Epanech.	medium	.0690	.0472	.0713	.0754	.0756	.0830	.0739	.0488	.0816	.0845	.0933	.0944
9	Epanech.	large	.0633	.0458	.0714	.0748	.0757	.0819	.0672	.0467	.0819	.0842	.0954	.0949
9	Quartic	small	.1213	.0635	.0999	.1062	.1049	.1162	.1288	.0617	.1018	.1080	.1084	.1201
9	Quartic	medium	.0743	.0491	.0729	.0775	.0769	.0854	.0799	.0505	.0824	.0859	.0921	.0951
9	Quartic	large	.0634	.0455	.0707	.0741	.0748	.0812	.0673	.0464	.0809	.0833	.0941	.0938

$n = 100$, $\theta_0 = .5$, 10,000 Monte Carlo replications. All kernel arguments were scaled down by their respective standard deviation.

V CONCLUSION

This paper has used the notion of local differencing to construct an estimator for fixed-effect panel data models with nonparametric unobserved heterogeneity. In line with the traditional estimators, the fixed-effect 'ignorance' towards the models' unit-specific latent components that underlies first-differencing the data allows for the unobserved heterogeneity to take on very general forms. The resulting local GMM estimator was shown to be consistent and asymptotically normal. It complements the recent results on identification of random coefficients in linear panel data models with exogenous covariates. With some work, the approach can be extended to a more general class of nonlinear panel data models.

Several questions and extensions of potential interest immediately suggest themselves. A pertinent question from an applied point of view is how to best choose the smoothing parameters. Ideally, a data-driven selection method should be found that determines the bandwidth in an optimal manner. Indeed, the Monte Carlo results suggest the smoothing used to be of substantial importance in small sample. A larger study of the sensitivity of the local GMM estimator to the form of the incidental functions and the stochastic behavior of their arguments would also be desirable. Related to this comment, it appears of interest to categorize all the information content in the data in terms of nonlinear moment conditions under additional assumptions such as homoskedasticity. Such an exercise in the spirit of Ahn and Schmidt (1995) would allow for efficiency gains and, arguably, improved finite-sample performance. Given the increased attention to cross-sectional dependence in the panel data literature—see Sarafidis and Wansbeek (2010) for an up-to-date review—it might be worthwhile investigating the potential for local first-differencing in the presence of factor structures in the disturbance processes. It seems reasonable to expect that the GMM estimators for such problems can be modified in a manner that is analogues to what was discussed here.

A final remark related to incorporating unit-specific non-stationary effects, such as time trends or time dummies. These are important empirical phenomena. Clearly, local first-differencing would break down under such a scenario as these variables, when transformed, are bounded away from zero. This is no different than in the binary-choice setting of Honoré and Kyriazidou (2000). An interesting extension to the model considered here, therefore, would be to construct a hybrid version that combines local differencing with a parametric random-coefficient specification such as the ones studied in Chamberlain (1992) and Arellano and Bonhomme (2010).

REFERENCES

Ahn, S. C. and P. Schmidt. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68:5–27.

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

Arellano, M. (2003). *Panel Data Econometrics*. Advanced Texts in Econometrics. Oxford University Press.

Arellano, M. (2004). Modelling optimal instrumental variables for dynamic panel data models. Unpublished manuscript.

Arellano, M. and S. Bond. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58:277–297.

Arellano, M. and S. Bonhomme. (2010). Identifying distributional characteristics in random coefficients panel data models. Unpublished manuscript.

Arellano, M. and O. Bover. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68:29–51.

Browning, M., M. Ejrnæs, and J. Alvarez. (2009). Modelling income processes with lots of heterogeneity. *Review of Economic Studies*, forthcoming.

Chamberlain, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica*, 60:567–596.

Donald, S. G. and W. K. Newey. (2000). A jackknife interpretation of the continuous updating estimator. *Economics Letters*, 67:239–243.

Gozalo, P. and O. Linton. (2000). Local nonlinear least squares: Using parametric information in nonparametric regression. *Journal of Econometrics*, 99:63–106.

Hall, A. R. (2005). *Generalized Method of Moments*. Advanced Texts in Econometrics. Oxford University Press.

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

Hansen, L. P., J. Heaton, and A. Yaron. (1996). finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics*, 14:262–280.

Hastie, T. and R. Tibshirani. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 55:757–796.

Holtz-Eakin, D., W. K. Newey, and H. S. Rosen. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56:1371–1395.

Honoré, B. E. and E. Kyriazidou. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68:839–874.

Lewbel, A. (2007). A local generalized method of moments estimator. *Economics Letters*, 94:124–128.

Mundlack, Y. (1961). Empirical production function free of management bias. *Journal of Farm Economics*, 43:44–56.

Mundlack, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46:69–85.

Powell, J. L. (1987). Semiparametric estimation of bivariate latent variable models. Unpublished manuscript.

Sarafidis, V. and T. Wansbeek. (2010). Cross-sectional dependence in panel data analysis. Unpublished manuscript.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126:25–51.