FIRST-DIFFERENCING IN PANEL DATA MODELS WITH INCIDENTAL FUNCTIONS

Koen Jochmans

Sciences Po, Department of Economics, 28 rue des Saints-Pères, 75007 Paris, France
koen.jochmans@sciencespo.fr

First version: July 2010. This version: July 10, 2014

This note discusses a class of models for panel data that accommodate between-group heterogeneity that is allowed to exhibit positive within-group variance. Such a setup generalizes the traditional fixed-effect paradigm in which between-group heterogeneity is limited to univariate factors that act like constants within groups. Notable members of the class of models considered are nonlinear regression models with additive heterogeneity and multiplicative-error models suitable for non-negative limited dependent variables. The heterogeneity is modelled as a non-parametric nuisance function of covariates whose functional form is fixed within groups but is allowed to vary freely across groups. A simple approach to perform inference in such situations is based on local first-differencing of observations within a given group. This leads to moment conditions that, asymptotically, are free of nuisance functions. Conventional GMM procedures may then be readily applied. In particular, under suitable regularity conditions, such estimators are consistent and asymptotically normal, and asymptotically-valid inference can be performed using a plug-in estimator of the asymptotic variance.

Keywords: *local differencing, nuisance function, panel data.*

INTRODUCTION

The linear fixed-effect model is a cornerstone model in applied microeconometrics. The introduction of intercept terms that are heterogenous across units allows to control for various permanent differences between units that cannot be observed by the researcher. For example, in the seminal work of Mundlak (1961, 1978) the aim is to control for managerial ability in the estimation of production functions. With Cobb-Douglas technology, log-output of firm i at time j equals

$$y_{ij} = x'_{ij}\alpha_0 + a_{ij},$$

where x_{ij} represents log-input factors such as capital and labor, α_0 is the corresponding vector of elasticities, and a_{ij} is total factor productivity. The latter will typically be correlated with the inputs, rendering the ordinary least-squares estimator of α_0 inconsistent. To estimate the elasticities from within-group variation, total factor productivity is decomposed as $a_{ij} = \lambda_i + \varepsilon_{ij}$, where ε_{ij} is assumed to be orthogonal to the production inputs but λ_i can be correlated with the x_{ij}. In this case, a within-group transformation will sweep out λ_i, after which least-squares can be applied to estimate α_0. The inclusion of fixed effects in this manner has become standard practice in applied work.

However, there are good reasons to believe that unobserved heterogeneity goes beyond what can be captured by such location parameters. In the production-function example, it seems natural that managerial ability depends on such things as experience, education, and sector-specific characteristics. As such, ability itself is the outcome of a production process, and it may be difficult to justify that it remains constant over the sampling period. A more appropriate way to control for managerial ability then could have $a_{ij} = \theta_i(v_{ij}) + \varepsilon_{ij}$ for some latent function θ_i that maps drivers v_{ij} such as experience and schooling into ability. In the same vain, in matching models, a_{ij} could represent the match-efficiency parameter. In the context of the labor market, Sedláček (2014) finds empirical evidence that matching efficiency is procyclical and is, at least partially, driven by the hiring standards of firms. Moreover, the matching literature has argued that the efficiency parameter should be endogenous to the agents' optimization behavior rather than exogenously determined.

This note suggests a simple way to conduct inference on common parameters in panel-data models with nonparametric incidental functions. Besides the linear setup just described, the approach can equally be used for models with multiplicative errors, such as models for count data, and for multinomial logit models, for example. In either case, staying true to the fixed-effect paradigm, the aim is to estimate a finite-dimensional parameter while controlling for between-group heterogeneity in a nonparametric manner. The difference with the traditional fixed-effect view, however, is that the heterogeneity is allowed to vary both within and between groups. This view on unobserved heterogeneity is different from the one taken in recent work on the linear random-coefficient model (Arellano and Bonhomme 2012; Graham and Powell 2012) and, as such, can serve as a useful complement.

1. LOCAL FIRST DIFFERENCING

1.1. Incidental functions

Consider a panel dataset consisting of two observations on n units. Restricting attention to two observations is without loss of generality. We let $y_i \equiv (y_{i1}, y_{i2})$ denote the outcome variables for unit i, and let $x_i \equiv (x_{i1}, x_{i2})$ and $v_i \equiv (v_{i1}, v_{i2})$ denote observable covariates. The distinction between the variables x_i and v_i will become clear below.

The workhorse fixed-effect model specifies unit i's response function as a linear function with a unit-specific intercept, as in

$$y_{ij} = \lambda_i + x_{ij}'\alpha_0 + \varepsilon_{ij} \tag{1.1}$$

for noise terms ε_{ij} and vector of slope coefficients α_0. Applications of this model are widespread. When $\mathbb{E}[\varepsilon_{ij}|x_i, \lambda_i] = 0$, an ordinary least-squares regression of $\Delta y_i \equiv y_{i2} - y_{i1}$ on $\Delta x_i \equiv x_{i2} - x_{i1}$ is known to yield a consistent point estimator of α_0 as $n \to +\infty$. Indeed,

$$\mathbb{E}[\Delta x_i \, (\Delta y_i - \Delta x_i'\alpha)] = 0$$

globally identifies α_0 provided $\mathbb{E}[\Delta x_i \Delta x_i']$ has full rank. When the covariates are not strictly exogenous, the above moment condition can be replaced by $\mathbb{E}[z_i \, (\Delta y_i - \Delta x_i'\alpha)] = 0$ for a vector of instrumental variables z_i. A leading case would be a dynamic model where $x_{ij} = y_{ij-1}$ and z_i contains further lags of the outcome variable; see, e.g., Arellano and Bond (1991).

In (1.1), α_0 is the parameter of interest. The traditional way of controlling for additional heterogeneity among agents is by introducing a set of strictly-exogenous control variables, v_{ij}, as additional regressors. This delivers a specification of the form

$$y_{ij} = \lambda_i + x_{ij}'\alpha_0 + v_{ij}'\beta_0 + \varepsilon_{ij}, \tag{1.2}$$

say. Here, the v_{ij} can be flexible polynomials specifications or other nonlinear transformations of the controls and, of course, may include interactions with x_{ij}. The choice of functional form is up to the researcher, and linearity is popular due to the resulting ease of computation via multiple regression. An approach that would prevent functional-form misspecification in the effect of the control variables would be to work with the partially-linear model

$$y_{ij} = \lambda_i + x_{ij}'\alpha_0 + \theta(v_{ij}) + \varepsilon_{ij}, \tag{1.3}$$

of which (1.2) is merely a special case. This is the approach advocated in the work of Robinson (1988). While he worked in a cross-sectional framework, it is quite obvious that his results can be extended to the panel-data version in (1.3). See Li and Stengos (1996), Ai, You, and Zhou (2014), and You and Zhou (2014) for a detailed analysis of such an approach in this type of model.

Nonetheless, a specification like (1.3) is less natural in a panel context than in a cross-sectional framework. Indeed, a main aim of the panel literature has been to devise flexible methods that allow for unobserved heterogeneity between units that stretches beyond what can be tackled with cross-section data. While (1.3) allows the impact of v_{ij} to be nonparametric, it is restricted to be identical across i. Recent empirical work has stressed the presence of excess heterogeneity across agents in microeconometric models. Guvenen (2009), Browning, Ejrnæs, and Alvarez (2010), Browning and Carro (2010), and Browning and Carro (2014), for example, provide extensive discussions and empirical evidence on this. An alternative extension of the Robinson framework that stays true to the fixed-effect tradition would be

$$y_{ij} = x_{ij}'\alpha_0 + \theta_i(v_{ij}) + \varepsilon_{ij}, \tag{1.4}$$

where, now, θ_i are unit-specific nonparametric functions, and the usual location parameter λ_i has been absorbed into it. A special case of (1.4) that has received some attention recently is the standard linear random-coefficient model (Swamy 1970; Chamberlain 1992b; Arellano and Bonhomme 2012). Another is the varying-coefficient model (Hastie and Tibshirani 1993). Nonetheless, the motivation for allowing for excess heterogeneity is clearly different in these cases.

A complication with (1.4), as opposed to (1.3), is that α_0 can no longer be identified through the approach of Robinson (1988). Indeed, an extension of his argument would require that $\mathbb{E}[\Delta y_i|v_i,\theta_i]$ and $\mathbb{E}[\Delta x_i|v_i,\theta_i]$ can be consistently estimated. Clearly, this is not possible under asymptotics where the number of observations per unit is held fixed. However, if $|\theta_i(v_{i2}) - \theta_i(v_{i1})| \leq \Theta_i(v_{i1}, v_{i2})\, \|\Delta v_i\|$ for some function Θ_i for which the expectation $\mathbb{E}[\Theta_i(v_{i1}, v_{i2})|\Delta v_i = v]$ exists for all v in a neighborhood of zero, then, provided that $\mathbb{E}[\varepsilon_{ij}|x_i, v_i, \theta_i]$ is a constant,

$$\mathbb{E}[\Delta x_i(\Delta y_i - \Delta x_i'\alpha_0)|\Delta v_i = 0] = 0$$

globally identifies α_0 if $\mathbb{E}[\Delta x_i \Delta x_i'|\Delta v_i = 0]$ has full rank. Indeed,

$$\alpha_0 = \mathbb{E}[\Delta x_i \Delta x_i'|\Delta v_i = 0]^{-1}\, \mathbb{E}[\Delta x_i \Delta y_i|\Delta v_i = 0]$$

under this condition. The smoothness condition on θ_i is fairly weak. Suppose that θ_i is continuously differentiable. Then its derivative, say θ_i', is locally bounded. Hence, $\Theta(v_{i1}, v_{i2}) = \sup_v |\theta_i'(v)|$ with the v restricted to the neighborhood $[\min\{v_{i1}, v_{i2}\}, \max\{v_{i1}, v_{i2}\}]$ satisfies the required Lipschitz-type smoothness condition. When the support of the v_{ij} is discrete, we need that $P(\Delta v_i = 0) > 0$. An estimator of α_0 would be

$$\alpha_n = \left(\frac{1}{n}\sum_{i=1}^n \Delta x_i \Delta x_i'\, \omega_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^n \Delta x_i \Delta y_i\, \omega_i\right),$$

where $\omega_i \equiv 1\{\Delta v_i = 0\}$. This estimator is $\sqrt{n}$-consistent and asymptotically normal under standard moment assumptions. When v_{ij} is continuous, the event $\{\Delta v_i = 0\}$ has probability zero and $\sqrt{n}$-consistent estimation will not be possible. However, under suitable regularity conditions, we can still perform asymptotically-valid inference on α_0 via α_n on redefining ω_i as

$$\omega_i = \frac{1}{h_n^{\dim v}}\, k\left(\frac{\Delta v_i}{h_n}\right)$$

for a chosen kernel function k and a bandwidth h_n.[1,2] Here, the convergence rate of α_n will be reduced to $\sqrt{nh_n^{\dim v}}$. We provide regularity conditions and more detailed asymptotic theory below. In either case, the approach consists of simply constructing ω_i for each i and then performing a weighted least-squares regression of Δy_i on Δx_i with weight ω_i. This estimator is similar in spirit to the one considered for sample-selection models by Kyriazidou (1997, 2001).

The θ_i can be seen as incidental functions, as opposed to the incidental parameters λ_i in the conventional setup in (1.1). Furthermore, the θ_i can be seen as draws from a distribution that depends on (x_i, v_i) but which is left unspecified. The approach just described does not estimate these functions but, rather, differences them out by focusing on the population of 'stayers' (Chamberlain 1984), that is, on units for which Δv_i lies in a shrinking neighborhood of zero. As such, this approach could be called local first differencing. Of course, a prerequisite to identification is that the support of v_{i1} and the support of v_{i2} cannot be disjoint. The leading example where this requirement would be violated is when the v_{ij} include time dummies or time trends. Such aggregate time effects are commonly used in applied work. Of course, they can easily be included in the traditional way, that is, by including them in a linear fashion and assigning them homogenous coefficients.

The use of stayers to recover parameters of interest from short panel data has recently also been used by Hoderlein and White (2012). They consider fully nonparametric structures of the form

$$y_{ij} = \theta_i(v_{ij}, \varepsilon_{ij}),$$

say, and study conditions under which local-average response functions can be identified and estimated. More precisely, they give conditions under which

$$\left.\frac{\partial\, \mathbb{E}[\Delta y_i | v_{i1}, \Delta v_i = v]}{\partial v}\right|_{v=0} = \mathbb{E}\left[\left. \frac{\partial \theta_i(v_{i1}, \varepsilon_{i1})}{\partial v_{i1}} \right| v_{i1}, \Delta v_i = 0 \right],$$

which is an average partial-effect for the subpopulation of stayers. Our setup is more modest in terms of generality and focuses on different parameters of interest. As such, we can allow x_{ij} to be predetermined as opposed to strictly exogenous and can accommodate discrete components in both x_{ij} and v_{ij}. Nonetheless, like in Hoderlein and White (2012) and Arellano and Bonhomme (2012), allowing for feedback toward the v_{ij} is complicated, as the distribution of the transitory shocks, ε_{ij}, may change after conditioning on the event $\Delta v_i = 0$.

1.2. Nonlinear specifications

The applicability of local first-differencing is not limited to the linear model. Indeed, any fixed-effect model where heterogenous intercepts can be accommodated can be extended to allow for incidental functions. The literature on panel data models is large, and we will not attempt to give a complete overview here. A rather exhaustive survey is provided by Arellano and Honoré (2001).

One obvious generalization would be to allow for a nonlinear relationship between y_{ij} and x_{ij} but to maintain additivity of the incidental function, as in

$$y_{ij} = \mu(x_{ij}; \alpha_0) + \theta_i(v_{ij}) + \varepsilon_{ij}, \qquad \mathbb{E}[\varepsilon_{ij} | x_i, v_i, \theta_i] = 0,$$

[1] As in standard nonparametric-regression theory, the choice of k has a much smaller impact than does the choice of h_n. An automated approach to selecting the bandwidth is to estimate it jointly with α_0, as in Härdle, Hall, and Ichimura (1993). See Appendix B for details and simulation experiments.

[2] Of course, both discrete and continuous variables can equally be accommodated by specifying kernel weights for the continuous elements of Δv_i and indicator functions for the discrete elements.

for some function μ that is known up to the Euclidean parameter α_0. Another type of nonlinearity that has proved important in panel data applications features in models of the form

$$y_{ij} = \mu(x_{ij};\alpha_0)\,\theta_i(v_{ij})\,\varepsilon_{ij}, \qquad \mathbb{E}[\varepsilon_{ij}|x_i, v_i, \theta_i] = 1.$$

A leading example of such a multiplicative model would be an exponential regression model with mean $\mathbb{E}[y_{ij}|x_i, v_i, \theta_i] = \exp\{\log\theta_i(v_{ij}) + x'_{ij}\alpha_0\}$. Here,

$$\left|\mathbb{E}\left[\frac{y_{i2}}{\mu(x_{i2};\alpha_0)} - \frac{y_{i1}}{\mu(x_{i1};\alpha_0)}\middle| x_i, v_i, \theta_i\right]\right| = |\theta_i(v_{i2}) - \theta_i(v_{i1})| \leq \Theta(v_{i1}, v_{i2})\,\|\Delta v_i\| \overset{\|\Delta v_i\|\downarrow 0}{\longrightarrow} 0.$$

In the conventional setup, fixed-effect estimation of multiplicative models of this form was discussed by Chamberlain (1992a) and Wooldridge (1997). Dynamic versions of this model can equally be handled; see Blundell, Griffith, and Windmeijer (2002).

The multinomial logit model with fixed effects is the prime example of the success of conditional maximum likelihood in panel models (Chamberlain 1980). A binary-choice version of a specification with incidental functions would have

$$y_{ij} = 1\{x'_{ij}\alpha_0 + \theta_i(v_{ij}) \geq \varepsilon_{ij}\}, \qquad F(e) \equiv P(\varepsilon_{ij} \leq e) = \frac{1}{1+\exp(-e)},$$

with the (x_{ij}, v_{ij}) independent of the ε_{ij}. An application of the conditional-likelihood argument shows that

$$\mathbb{E}[1\{\Delta y_i = 1\} - F(\Delta x'_i\alpha_0)|x_i, v_i, \Delta y_i \neq 0, \Delta v_i = 0] = 0,$$

which is free of θ_i. The optimal unconditional moment condition in the sense of Chamberlain (1992b) equals

$$\mathbb{E}[\Delta x_i\,(1\{\Delta y_i = 1\} - F(\Delta x'_i\alpha_0))\,1\{\Delta y_i \neq 0\}|\Delta v_i = 0] = 0$$

and can be seen as the first-order condition associated with a local conditional likelihood. It is useful to note that this moment condition is very similar to the first-order condition of the estimator of Honoré and Kyriazidou (2000) for a dynamic logit model with exogenous regressors.

In each of the examples just mentioned, it is easy to construct a GMM estimator in which the usual moment condition is complemented with the kernel weight ω_i as described above. We will provide asymptotic theory in the next section.

There are several other models that could be extended to allow for incidental functions. Some interesting examples are truncated- and censored regression models (Honoré 1992), as well as general transformation models and generalized-regression models (Abrevaya 1999, 2000). The resulting estimators would have similar asymptotic properties. However, they are M-estimators rather than GMM estimators, and the associated criterion functions are characterized by a certain degree of non-smoothness. As such, they will not fit exactly the generic setup entertained below.

2. ASYMPTOTIC THEORY

Consider a generic setup in which a Euclidean parameter $\alpha_0 \in \mathcal{A}$ is identified through the moment condition

$$\mathbb{E}[m(y_i, x_i;\alpha_0)|\Delta v_i = 0] = 0,$$

where m is a vector function that is known up to α_0. An empirical counterpart to the population moment at α is

$$\sigma_n(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{m(y_i, x_i;\alpha)}{h_n^{\dim v}}\,k\left(\frac{\Delta v_i}{h_n}\right),$$

where h_n is a non-negative bandwidth sequence that is $o(1)$ and k is a kernel function. Regularity conditions on h_n and k are collected in Assumption 3 below. A GMM estimator of α_0 based on $\sigma_n(\alpha)$ is then given by

$$\alpha_n \equiv \arg\min_{\alpha \in \mathcal{A}} \sigma_n(\alpha)' W_n \, \sigma_n(\alpha),$$

where W_n denotes a given positive-definite weight matrix. This section provides distribution theory for α_n in the form of a consistency result and an asymptotic-normality result. The proofs are given in Appendix A.

Some elementary regularity conditions are collected in Assumption 1.

Assumption 1. *$\mathcal{A}$ is a compact set and α_0 is interior to it. m is twice continuously differentiable in α with derivatives m' and m''. The distribution of Δv_i is absolutely continuous and the associated density function is strictly positive in a neighborhood of zero.*

Let $\|\cdot\|$ denote the Euclidean and Frobenius norms. To state sufficient conditions for consistency, let

$$\sigma(v;\alpha) \equiv \mathbb{E}[m(y_i, x_i; \alpha)|\Delta v_i = v] \, f(v),$$

for f the density of Δv_i.

Assumption 2. *For all $\alpha \in \mathcal{A}$, $\mathbb{E}[\|m(y_i, x_i; \alpha)\|^2]$ and $\mathbb{E}[\|m'(y_i, x_i; \alpha)\|^2]$ are finite, $\|\sigma(v;\alpha)\|$ is bounded in v, and $\sigma(v;\alpha)$ is continuous in v in a neighborhood of zero.*

Assumption 3. *$k : \mathcal{R}^{\dim v} \rightarrow \mathcal{R}$ is a bounded and symmetric sth-order kernel function.*

The conditions in Assumptions 2 and 3 are conventional. We refer to Li and Racine (2007) for a definition, examples, and discussion on kernel functions that satisfy Assumption 3.

The consistency result is stated in Theorem 1.

Theorem 1. *Let Assumptions 1–3 hold. Suppose that $\|W_n - W_0\| = o_P(1)$ for W_0 non-stochastic and positive definite. Then $\|\alpha_n - \alpha_0\| = o_P(1)$.*

To derive the limit distribution of α_n we need an additional set of conditions. We let

$$\Sigma(v;\alpha) = \frac{\partial \sigma(v;\alpha)}{\partial \alpha'}, \qquad \Delta(v;\alpha) = \mathbb{E}[m(y_i, x_i; \alpha) \, m(y_i, x_i; \alpha)'|\Delta v_i = v] \, f(v)$$

in the following assumption.

Assumption 4. *For all $\alpha \in \mathcal{A}$, $\mathbb{E}[\|m''(y_i, x_i, \alpha)\|^2]$ is finite, $\|\Sigma(v, \alpha)\|$ is bounded in v, and $\Sigma(v, \alpha)$ is continuous in v in a neighborhood of zero. $\mathbb{E}[\|m(y_i, x_i; \alpha_0)\|^3|\Delta v_i = v] \, f(v)$ is bounded. $\Delta(v;\alpha_0)$ is continuous in v in a neighborhood of zero and $\|\Delta(v;\alpha_0)\|$ is bounded. $\sigma(v;\alpha_0)$ is s-times continuously-differentiable with bounded derivatives.*

Let $\Sigma \equiv \Sigma(0, \alpha_0)$ and $\Delta \equiv \Delta(0, \alpha_0) \int_{-\infty}^{+\infty} k(\eta)^2 \, \mathrm{d}\eta$. Theorem 2 gives the asymptotic distribution of α_n.

Theorem 2. *Let Assumptions 1–4 hold. Suppose that Σ has maximal column rank, that Δ is positive definite, and that $\|W_n - W_0\| = o_P(1)$ for W_0 non-stochastic and positive definite. Then*

$$\sqrt{n h_n^{\dim v}} (\alpha_n - \alpha_0) \overset{A}{\sim} \mathcal{N}\left(0, (\Sigma' W_0 \Sigma)^{-1} \, (\Sigma' W_0 \Delta W_0 \Sigma) \, (\Sigma' W_0 \Sigma)^{-1}\right)$$

provided $\sqrt{n h_n^{\dim v}} \rightarrow +\infty$ and $\sqrt{n h_n^{\dim v}} h_n^s \rightarrow 0$.

The matrices Σ and Δ are estimated consistently by

$$\Sigma_n \equiv \frac{1}{n}\sum_{i=1}^n \frac{m'(y_i,x_i;\alpha_n)}{h_n^{\dim v}}\, k\left(\frac{\Delta v_i}{h_n}\right), \qquad \Delta_n \equiv \frac{1}{n}\sum_{i=1}^n \frac{m(y_i,x_i;\alpha_n)\, m(y_i,x_i;\alpha_n)'}{h_n^{\dim v}}\, k\left(\frac{\Delta v_i}{h_n}\right)^2,$$

respectively

ACKNOWLEDGMENTS

I am grateful to Jaap Abbring and three referees, Manuel Arellano, and Stefan Hoderlein.

APPENDIX A: PROOFS OF THEOREMS

Proof of Theorem 1. Let $\sigma(\alpha) \equiv \sigma(0,\alpha)$. Given identification, the regularity conditions in Assumption 1, and the fact that $W_n \xrightarrow{P} W_0$, we only need to verify

$$\sup_{\alpha\in\mathcal{A}} \|\sigma_n(\alpha) - \sigma(\alpha)\| = o_P(1)$$

to establish consistency; see Theorem 2.1 of Newey and McFadden (1994). Because m is differentiable, $\sup_{\alpha\in\mathcal{A}} \mathbb{E}[\|m'(y_i,x_i;\alpha)\|^2]$ is finite, and k is bounded, Lemma 2.9 in Newey and McFadden (1994) further states that that it suffices to prove that $\|\sigma_n(\alpha)-\sigma(\alpha)\| = o_P(1)$ for all $\alpha\in\mathcal{A}$. Fix $\alpha\in\mathcal{A}$. By the triangle inequality,

$$\|\sigma_n(\alpha)-\sigma(\alpha)\| \leq \|\sigma_n(\alpha)-\mathbb{E}[\sigma_n(\alpha)]\| + \|\mathbb{E}[\sigma_n(\alpha)]-\sigma(\alpha)\|.$$

Assumption 2 and Assumption 3 imply that $\|\sigma_n(\alpha)-\mathbb{E}[\sigma_n(\alpha)]\| = o_p(1)$ by the law of large numbers. Dominated convergence implies that

$$\mathbb{E}[\sigma_n(\alpha)] = \int_{-\infty}^{+\infty} \frac{\sigma(v;\alpha)}{h_n^{\dim v}}\, k\left(\frac{\Delta v}{h_n}\right) \mathrm{d}v = \int_{-\infty}^{+\infty} \sigma(h_n\eta;\alpha)\, k(\eta)\mathrm{d}\eta \to \sigma(\alpha),$$

and so $\|\mathbb{E}[\sigma_n(\alpha)]-\sigma(\alpha)\| = o_P(1)$. Thus, $\|\sigma_n(\alpha)-\sigma(\alpha)\| = o_P(1)$. This holds for any $\alpha\in\mathcal{A}$, and so consistency has been shown. □

Proof of Theorem 2. We will show (i) $\sqrt{nh_n^{\dim v}}\sigma_n(\alpha_0) \overset{A}{\sim} \mathcal{N}(0,\Delta)$ and (ii) $\sup_{\alpha\in\mathcal{A}}\|\Sigma_n(\alpha)-\Sigma(\alpha)\| = o_P(1)$. The asymptotic distribution of the estimator then follows from the linearization

$$\sqrt{nh_n^{\dim v}}(\alpha_n-\alpha_0) = -(\Sigma' W_0 \Sigma)^{-1}\, \Sigma' W_0 \sqrt{nh_n^{\dim v}}\sigma_n(\alpha_0) + o_P(1)$$

by an application of the delta method. To show (i), first observe that

$$\sqrt{nh_n^{\dim v}}\sigma_n(\alpha_0) = \sqrt{nh_n^{\dim v}}\big(\sigma_n(\alpha_0) - \mathbb{E}[\sigma_n(\alpha_0)]\big) + \sqrt{nh_n^{\dim v}}\, \mathbb{E}[\sigma_n(\alpha_0)].$$

The second term on the right-hand side is a bias term. By an sth-order expansion and Assumptions 3 and 4,

$$\mathbb{E}[\sigma_n(\alpha_0)] = \int_{-\infty}^{+\infty} \sigma(h_n\eta;\alpha_0)\, k(\eta)\, \mathrm{d}\eta = O(h_n^s).$$

As $\sqrt{nh_n^{\dim v}}h_n^s = o(1)$, $\sqrt{nh_n^{\dim v}}\, \mathbb{E}[\sigma_n(\alpha_0)] = o(1)$ and the bias term is asymptotically negligible. The leading term satisfies the conditions of Lyapunov's central limit theorem. To see this, write

$$\sqrt{nh_n^{\dim v}}\big(\sigma_n(\alpha_0) - \mathbb{E}[\sigma_n(\alpha_0)]\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_i - \mathbb{E}[\gamma_i], \qquad \gamma_i \equiv \frac{m(y_i,x_i;\alpha_0)}{\sqrt{h_n^{\dim v}}}\, k\left(\frac{\Delta v_i}{h_n}\right).$$

Then $\mathbb{E}[\gamma_i] = o(1)$ and

$$\mathrm{var}[\gamma_i] = \mathbb{E}[\gamma_i\gamma_i'] - \mathbb{E}[\gamma_i]\, \mathbb{E}[\gamma_i'] = \int_{-\infty}^{+\infty} \frac{\Delta(v;\alpha_0)}{h_n^{\dim v}}\, k\left(\frac{\Delta v}{h_n}\right)^2 \mathrm{d}v + o(1) \to \Delta(0;\alpha_0) \int_{-\infty}^{+\infty} k(\eta)^2 \mathrm{d}\eta = \Delta,$$

by a bounded-convergence argument and Assumption 4. Finally, also Lyapunov's condition is satisfied, because

$$\sum_{i=1}^{n} \mathbb{E}\left[\left\|\frac{\gamma_i}{\sqrt{n}}\right\|^3\right] \leq \frac{1}{\sqrt{nh_n^{\dim v}}} \int_{-\infty}^{+\infty} \frac{\mathbb{E}[\|m(y_i,x_i;\alpha_0)\|^3|\Delta v_i = v]\, f(v)}{h_n^{\dim v}} \left|k\left(\frac{v}{h_n}\right)\right|^3 \mathrm{d}v = O\left(\frac{1}{\sqrt{nh_n^{\dim v}}}\right),$$

which vanishes as $n \to +\infty$. This establishes (i). To verify (ii) one can proceed as in the proof of Theorem 1. In particular, Lemma 2.9 of Newey and McFadden (1994) may again be applied. By the moment conditions in Assumption 4 we have that $\|\Sigma_n(\alpha) - \mathbb{E}[\Sigma_n(\alpha)]\| = o_P(1)$. An application of the bounded convergence theorem similarly shows that $\mathbb{E}[\Sigma_n(\alpha)] \to \Sigma(\alpha)$. Uniform convergence of the Jacobian matrix follows and the proof is complete. □

APPENDIX B: SIMULATIONS

The results from a small set of Monte Carlo experiments are collected as supplementary material. We consider two designs for two models. In the first design we draw $x_{i1} \sim \mathcal{N}(0,1)$ and $x_{i2} \sim \mathcal{N}(\frac{1}{2}x_{i1},1)$, and $v_{i1} \sim \mathcal{N}(0,1)$ and $v_{i2} \sim \mathcal{N}(\frac{1}{2}v_{i1},1)$, so that the x_{it} and v_{it} are independent. In the second design we induce dependence by generating $v_{i1} \sim \mathcal{N}(\frac{1}{2}x_{i1},1)$ and $v_{i2} \sim \mathcal{N}(\frac{1}{2}x_{i2}+\frac{1}{2}v_{i1},1)$. In either design, we draw

$$y_{it}^{*} = x_{it}\alpha_0 + v_{it},$$

and subsequently obtain the outcome variable as either (i) $y_{it} \sim \mathcal{N}(y_{it}^{*},1)$ (linear regression model) or (ii) $y_{it} \sim \text{Poisson}(\exp(y_{it}^{*}))$ (exponential regression model). In either case, we estimate α_0 from a single moment condition, with instrument Δx_i, and so we minimize $|\sigma_n(\alpha)|^2$. For implementation we use the fourth-order kernel $k(\eta) = (\frac{3}{2} - \frac{1}{2}\eta^2)\phi(\eta)$ with bandwidth $h_n = c_n\, n^{-1/7}$ for some constant c_n. The constant c_n is obtained in a data-driven manner, by minimizing the GMM objective function jointly with respect to α and c_n; see Härdle, Hall, and Ichimura (1993) for this proposal in a different context. Although we claim no optimality for the bandwidth-selection method just described, such an automatic procedure for selecting the bandwidth makes the implementation of our procedure quite straightforward. Table 1 reports the bias, standard deviation, and empirical rejection frequency of 95%-confidence intervals for our estimator for each of the designs and models described above. The results were obtained over 10,000 Monte Carlo replications.

Table 1. Simulation results for α_n

n	α_0	bias	std	size	bias	std	size
		linear model (i)					
		independence			dependence		
250	1	.0004	.1228	.0505	.0201	.1310	.0512
500	1	−.0015	.0933	.0507	.0139	.0990	.0512
1000	1	.0003	.0703	.0498	.0109	.0758	.0500
2500	1	−.0005	.0485	.0542	.0060	.0521	.0514
5000	1	.0000	.0359	.0496	.0044	.0387	.0521
		exponential model (ii)					
		independence			dependence		
250	1	.0295	.1655	.0457	.0378	.1414	.0450
500	1	.0180	.1181	.0464	.0209	.0983	.0425
1000	1	.0097	.0897	.0504	.0153	.0726	.0501
2500	1	.0055	.0611	.0507	.0079	.0479	.0514
5000	1	.0031	.0454	.0532	.0052	.0351	.0484

REFERENCES

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

Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics 95*, 1–23.

Ai, C., J. You, and Y. Zhou (2014). Estimation of fixed effects panel data partially linear additive regression models. *Econometrics Journal 17*, 83–106.

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 (2012). Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies 79*, 987–1020.

Arellano, M. and B. E. Honoré (2001). Panel data models: Some recent developments. In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume V, Chapter 53, pp. 3229–3329. Elsevier.

Blundell, R. W., R. Griffith, and F. Windmeijer (2002). Individual effects and dynamics in count data models. *Journal of Econometrics 108*, 113–131.

Browning, M. and J. M. Carro (2010). Heterogeneity in dynamic discrete choice models. *Econometrics Journal 13*, 1–39.

Browning, M. and J. M. Carro (2014). Dynamic binary outcome models with maximal heterogeneity. *Journal of Econometrics 178*, 805–823.

Browning, M., M. Ejrnæs, and J. Alvarez (2010). Modeling income processes with lots of heterogeneity. *Review of Economic Studies 77*, 1353–1381.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies 47*, 225–238.

Chamberlain, G. (1984). Panel data. In Z. Griliches and M. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, Chapter 22, pp. 1247–1315. Elsevier.

Chamberlain, G. (1992a). Comment: Sequential moment restrictions in panel data. *Journal of Business and Economic Statistics 10*, 20–26.

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

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*, 2105–2152.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics 12*, 58–79.

Härdle, W., P. Hall, and H. Ichimura (1993). Optimal smoothing in single-index models. *Annals of Statistics 21*, 157–178.

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

Hoderlein, S. and H. White (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics 168*, 300–314.

Honoré, B. E. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica 60*, 553–565.

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

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica 65*, 1335–1364.

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *Review of Economic Studies 68*, 543–572.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Li, Q. and T. Stengos (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics 71*, 389–397.

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

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

Newey, W. K. and D. L. McFadden (1994). Large sample estimation and hypothesis testing. In R. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Elsevier.

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica 56*, 931–954.

Sedláček, P. (2014). Match efficiency and firms' hiring standards. *Journal of Monetary Economics 62*, 123–133.

Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica 38*, 311–323.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory 13*, 667–678.

You, J. and X. Zhou (2014). Asymptotic theory in fixed effects panel data seemingly unrelated partially linear regression models. *Econometric Theory*, forthcoming.