Journal of Econometrics 185 (2015) 435–452

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Efficient inference on fractionally integrated panel data models with fixed effects

Peter M. Robinson [a,*], Carlos Velasco [b]

[a] London School of Economics, United Kingdom
[b] Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Article history:
Received 21 March 2013
Received in revised form
27 June 2014
Accepted 12 December 2014
Available online 31 December 2014

JEL classification:
C12
C13
C23

Keywords:
Panel data
Fractional time series
Estimation
Testing
Bias correction

ABSTRACT

A dynamic panel data model is considered that contains possibly stochastic individual components and a common stochastic time trend that allows for stationary and nonstationary long memory and general parametric short memory. We propose four different ways of coping with the individual effects so as to estimate the parameters. Like models with autoregressive dynamics, ours nests $I(1)$ behaviour, but unlike the nonstandard asymptotics in the autoregressive case, estimates of the fractional parameter can be asymptotically normal. For three of the estimates, establishing this property is made difficult due to bias caused by the individual effects, or by the consequences of eliminating them, which appears in the central limit theorem except under stringent conditions on the growth of the cross-sectional size N relative to the time series length T, though in case of two estimates these can be relaxed by bias correction, where the biases depend only on the parameters describing autocorrelation. For the fourth estimate, there is no bias problem, and no restrictions on N. Implications for hypothesis testing and interval estimation are discussed, with central limit theorems for feasibly bias-corrected estimates included. A Monte Carlo study of finite-sample performance is included.

1. Introduction

Important features of many econometric models for panel data are unobserved individual fixed effects and temporal dynamics that possibly allow for nonstationarity. When the cross-sectional dimension is large the individual effects cause an incidental parameters problem that heavily determines methodology, which has been predominately developed in the context of autoregressive, including possibly unit root, dynamics. A recent textbook treatment is Hsiao (2014). The present paper focuses on the incidental parameters problem in the context of fractional dynamics, which offer some advantages over autoregressions. A simple model for an observable array $\{y_{it}\}$ is

$$\lambda_t (L; \theta_0) (y_{it} - \alpha_i) = \varepsilon_{it}, \tag{1}$$

for $i = 1, \ldots, N$, $t = 0, 1, \ldots, T$. The unobserved individual effects $\{\alpha_i, i \geq 1\}$ are subject to little, if any, more detailed specification in the sequel; the unobserved innovations $\{\varepsilon_{it}, i \geq 1, t \geq 0\}$ are throughout assumed to be independent and identically distributed (iid) and to satisfy $E\varepsilon_{it} = 0, E\varepsilon_{it}^4 < \infty$; θ_0 is a $(p+1)\times 1$ parameter vector, known only to lie in a given compact subset Θ of R^{p+1}; L is the lag operator; for any $\theta \in \Theta$ and each $t \geq 0$,

$$\lambda_t (L; \theta) = \sum_{j=0}^{t} \lambda_j (\theta) L^j \tag{2}$$

truncates the expansion

$$\lambda (L; \theta) = \sum_{j=0}^{\infty} \lambda_j (\theta) L^j,$$

where the $\lambda_j (\theta)$ are given functions. We are concerned with $\lambda (L; \theta)$ having the particular structure

$$\lambda (L; \theta) = \Delta^{\delta} \psi (L; \xi),$$

where δ is a scalar, ξ is a $p \times 1$ vector, $\theta = (\delta, \xi')'$, the prime denoting transposition, and the functions Δ^{δ} and $\psi (L; \xi)$ are described as follows. With $\Delta = 1 - L$, Δ^{δ} has the expansion

$$\Delta^{\delta} = \sum_{j=0}^{\infty} \pi_j (\delta) L^j, \qquad \pi_j (\delta) = \frac{\Gamma (j - \delta)}{\Gamma(-\delta)\Gamma(j+1)},$$

for non-integer $\delta > 0$, while for integer $\delta = 0, 1, \ldots, \pi_j (\delta) = 1(j = 0, 1, \ldots, \delta) (-1)^j \delta (\delta - 1) \cdots (\delta - j + 1) / j!$, taking $0/0 =$

* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
E-mail address: p.m.robinson@lse.ac.uk (P.M. Robinson).

http://dx.doi.org/10.1016/j.jeconom.2014.12.003

1 and 1 (.) to be the indicator function; $\psi(L;\xi)$ is a known function of its arguments such that for complex-valued x, $|\psi(x;\xi)| \neq 0$, $|x| \leq 1$ and in the expansion

$$\psi(L;\xi) = \sum_{j=0}^{\infty} \psi_j(\xi) L^j,$$

the coefficients $\psi_j(\xi)$ satisfy

$$\psi_0(\xi) = 1, \qquad \psi_j(\xi) = O(\exp(-c(\xi) j)), \tag{3}$$

where $c(\xi)$ is a positive-valued function of ψ. Note that

$$\lambda_j(\theta) = \sum_{k=0}^{j} \pi_{j-k}(\delta)\psi_k(\xi), \quad j \geq 0. \tag{4}$$

The fractional operator Δ^{δ} bestows possible stationary (when $0 < \delta < 1/2$) or nonstationary (when $\delta \geq 1/2$) long memory on $y_{it} - \alpha_i$, while $\psi(L;\xi)$ adds possible short memory structure, for example representing the autoregressive operator of a stationary and invertible autoregressive moving average process with combined order p, or of an exponential spectrum model (Bloomfield (1973)). The truncation in (2) is motivated mainly by a desire to allow for $\delta \geq 1/2$, when $\Delta^{-\delta}$, and thus $\lambda^{-1}(L;\theta)$, do not converge. On the other hand we can write (1) as

$$y_{it} = \alpha_i + \lambda_t^{-1}(L;\theta_0)\varepsilon_{it} = \alpha_i + \lambda^{-1}(L;\theta_0)\{\varepsilon_{it} 1(t \geq 0)\}.$$

It is possible that ξ_0 is empty, i.e. $p = 0$ and $\psi(x;\xi) \equiv 1$ *a priori*, in which case for each i, $y_{it} - \alpha_i$ has pure fractional dynamics. Our interest is in statistical inference on $\theta_0 = (\delta_0, \xi_0')'$, and especially on δ_0 with ξ_0 regarded as a nuisance parameter.

For each i we can call $y_{it} - \alpha_i$ an $I(\delta_0)$ process. Temporarily taking $\psi(x;\xi) \equiv 1$ for simplicity, we can write

$$y_{it} = \alpha_i + \sum_{j=0}^{t} \pi_j(-\delta_0)\varepsilon_{i,t-j}, \tag{5}$$

whence when $\delta_0 = 1$,

$$y_{it} = \alpha_i + \sum_{j=0}^{t} \varepsilon_{i,t-j}. \tag{6}$$

The latter results also on taking $\rho = 1$ in the autoregressive scheme popular in the dynamic panel data literature:

$$y_{it} = \alpha_i + \sum_{j=0}^{t} \rho^j \varepsilon_{i,t-j}. \tag{7}$$

The typical alternatives to $\rho = 1$ covered by (7) are the stationary ones $\rho \in (-1, 1)$ or the explosive ones $\rho > 1$. Other versions of the autoregressive panel data model are

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \varepsilon_{it}, \quad t > 0, \tag{8}$$

and

$$y_{it} = \alpha_i + u_{it}, \quad u_{it} = \rho u_{i,t-1} + \varepsilon_{it},\ t > 0, \tag{9}$$

with $\rho \in (-1, 1]$; note that (9) implies that

$$y_{it} = (1-\rho)\alpha_i + \rho y_{i,t-1} + \varepsilon_{it}, \quad t > 0,$$

so that α_i is eliminated when $\rho = 1$. The usual aim in (7), (8) or (9) is estimating ρ or unit root testing. As one recent reference, Han and Phillips (2010) develop inference based on generalized method-of-moment estimates. Note that in the fractional model (5), the weights $\pi_j(-\delta_0)$ have decay or growth that is, unlike in (7), not exponential but algebraic, since, for any δ,

$$\pi_j(\delta) = \frac{1}{\Gamma(-\delta)} j^{-\delta-1}(1 + O(j^{-1})) \quad \text{as } j \to \infty. \tag{10}$$

The moving average weights in the more general model (1) have the same rate, in particular, by (4) and summation-by-parts,

$$\begin{aligned}
\lambda_j(\theta) &= \sum_{k=0}^{j-1}\left(\pi_{j-k}(\delta) - \pi_{j-k-1}(\delta)\right)\sum_{l=0}^{k}\psi_l(\xi) + \sum_{l=0}^{j}\psi_l(\xi) \\
&= \psi(1;\xi)\sum_{k=0}^{j-1}\left(\pi_{j-k}(\delta) - \pi_{j-k-1}(\delta)\right) + \psi(1;\xi) \\
&\quad - \sum_{k=0}^{j-1}\left(\pi_{j-k}(\delta) - \pi_{j-k-1}(\delta)\right)\sum_{l=k+1}^{\infty}\psi_l(\xi) - \sum_{l=j+1}^{\infty}\psi_l(\xi) \\
&= \psi(1;\xi)\pi_j(\delta) \\
&\quad + O\left(\sum_{k=0}^{j-1}(j-k)^{-\delta-2}\exp(-c(\xi)k) + \exp(-c(\xi)j)\right) \\
&= \frac{\psi(1;\xi)}{\Gamma(-\delta)} j^{-\delta-1}(1 + O(j^{-1})) \quad \text{as } j \to \infty,
\end{aligned} \tag{11}$$

using (10) and (3), where we note that the exponential decay requirement in the latter ensures that (11) holds for all $\delta > 0$.

As is well known from the time series literature the fractional class described by $\lambda_t(L;\theta_0)$ has a smoothness at $\delta_0 = 1$ (and elsewhere) that the autoregressive class lacks. A consequence established in that literature is that large sample inference based on an approximate Gaussian pseudo likelihood can be expected to entail standard limit distribution theory; in particular, Lagrange multiplier tests on θ_0 (for example of the $I(1)$ hypothesis $\delta_0 = 1$) are asymptotically χ^2 distributed with classical local power properties, and estimates of θ_0 are asymptotically normally distributed with the usual parametric rate (see Robinson (1991, 1994), Beran (1995), Velasco and Robinson (2000), Hualde and Robinson (2011)). This is the case whether δ_0 lies in the stationary region $(0, 1/2)$ or the nonstationary one $[1/2, \infty)$ (or, also, the negative dependent region $(-\infty, 0)$).

If N is regarded as fixed while $T \to \infty$, (1) is just a multivariate fractional model, with a vector, possibly stochastic, location. But in many practical applications N is large, and even when smaller than T, is more reasonably treated as diverging in asymptotic theory if T is. In that case inference on θ_0 is considerably complicated by an incidental parameters problem. In this paper we present and justify several approaches that resolve this question. We throughout employ asymptotic theory with respect to T diverging, where either N increases with T or stays fixed, and both cases are covered by indexing with respect to T only. In (1) the interest is in estimating θ_0 (efficiently, perhaps with some *a priori* knowledge on the range of allowed values) and testing hypotheses such as $I(1)$, $\delta_0 = 1$, or of absence of short memory structure, which might entail $\psi_0 = 0$. Hassler et al. (2011) have recently developed tests in a panel with a more general temporal dependence structure which is allowed to vary across units, and with allowance for cross-sectional dependence, but without allowing for individual effects and keeping N fixed as $T \to \infty$.

The following section introduces four rival estimates of θ_0. Three are versions of time series conditional-sum-of-squares (CSS) estimates, recently treated in a general fractionally integrated setting by Hualde and Robinson (2011), one of which ignores the fixed effects, while the other two correct for them by regression and first differencing, respectively. The fourth is a Gaussian pseudo-maximum likelihood estimate (PMLE) based on the differenced model, and is somewhat more onerous computationally. Section 3 contains consistency theorems. In Section 4 the estimates are shown to be asymptotically normal. For the 3 CSS estimates, unless the restriction on the growth of N relative to T is very stringent, asymptotic biases in the central limit theorem are present,