Journal of Econometrics 175 (2013) 35–45

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

First difference maximum likelihood and dynamic panel estimation☆

Chirok Han[a,*], Peter C.B. Phillips[b,c,d,e]

[a] Korea University, Republic of Korea
[b] Yale University, United States
[c] University of Auckland, New Zealand
[d] University of Southampton, United Kingdom
[e] Singapore Management University, Singapore

ARTICLE INFO

Article history:
Received 3 June 2011
Received in revised form
11 February 2013
Accepted 11 March 2013
Available online 20 March 2013

JEL classification:
C22
C23

Keywords:
Asymptote
Bounded support
Dynamic panel
Efficiency
First difference MLE
Likelihood
Quartic equation
Restricted extremum estimator

ABSTRACT

First difference maximum likelihood (FDML) seems an attractive estimation methodology in dynamic panel data modeling because differencing eliminates fixed effects and, in the case of a unit root, differencing transforms the data to stationarity, thereby addressing both incidental parameter problems and the possible effects of nonstationarity. This paper draws attention to certain pathologies that arise in the use of FDML that have gone unnoticed in the literature and that affect both finite sample performance and asymptotics. FDML uses the Gaussian likelihood function for first differenced data and parameter estimation is based on the whole domain over which the log-likelihood is defined. However, extending the domain of the likelihood beyond the stationary region has certain consequences that have a major effect on finite sample and asymptotic performance. First, the extended likelihood is not the true likelihood even in the Gaussian case and it has a finite upper bound of definition. Second, it is often bimodal, and one of its peaks can be so peculiar that numerical maximization of the extended likelihood frequently fails to locate the global maximum. As a result of these pathologies, the FDML estimator is a restricted estimator, numerical implementation is not straightforward and asymptotics are hard to derive in cases where the peculiarity occurs with non-negligible probabilities. The peculiarities in the likelihood are found to be particularly marked in time series with a unit root. In this case, the asymptotic distribution of the FDMLE has bounded support and its density is infinite at the upper bound when the time series sample size $T \to \infty$. As the panel width $n \to \infty$ the pathology is removed and the limit theory is normal. This result applies even for T fixed and we present an expression for the asymptotic distribution which does not depend on the time dimension. We also show how this limit theory depends on the form of the extended likelihood.

1. Introduction

Maximum likelihood estimation based on first-differenced data (FDML) has recently attracted attention as an alternative estimation methodology to conventional maximum likelihood (ML) and GMM approaches in dynamic panel models (Hsiao et al., 2002; Kruiniger, 2008). FDML appears to offer certain immediate advantages in dynamic panels with fixed effects. Unlike unconditional ML where fixed effects are treated as parameters to estimate, FDML is free from the incidental parameter problem (Neyman and Scott, 1948) because nuisance individual effects have already been eliminated before deriving the likelihood. In addition, the differenced data are stationary whether the original data are stationary or integrated, and hence the presence of a unit root does not appear to require any special treatment or modification of the likelihood function. This feature is deemed especially useful when panel data show a large degree of persistence.

These advantages, coupled with the computational convenience of modern numerical optimization, have spurred the use of FDMLE in applied research. The empirical literature dates back to MaCurdy (1982). But there has been little research on the method's properties or on certain of its peculiarities such as negative variance estimates that are known to arise in its implementation by numerical optimization. Most importantly, it seems not to have been recognized in the literature that FDMLE is *not* a maximum likelihood procedure because the 'likelihood' that is used in optimization is based on analytically extending the stationary likelihood outside the stationary region. The resulting function is *not* a

☆ CH acknowledges support from the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-332-B00026). PCBP acknowledges support from the NSF under Grant No. SES09-56687.

* Correspondence to: Department of Economics, Korea University, Anam-dong, Seongbuk-gu, Seoul, Republic of Korea. Tel.: +82 2 3290 2205.
E-mail address: chirokhan@korea.ac.kr (C. Han).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.03.003

true likelihood outside the stationary region even though it is well defined for certain nonstationary regions. This feature of FDMLE is subtle, which partly explains why it has gone unnoticed in the literature for so long. But it has significant implications and leads to further complications, including an upper bound restriction on the domain that affects both finite sample theory and asymptotic behavior. An investigator may, of course, choose *a priori* to restrict the domain of the autoregressive roots to the unit circle, but in this event an appropriate asymptotic theory that accounts for the restriction would need to be used in practical work.

Wilson (1988) provided an exact likelihood for the differenced data generated from a stationary AR(1) process based on Ansley's (1979) expression for ARMA(1, 1), and discovered in simulations that FDMLE outperforms the maximum likelihood (ML) estimator in terms of mean squared error for small samples. Hsiao et al. (2002); hereafter HPT) studied FDML in linear dynamic panel models with wide short panels – that is panels with large cross sectional dimension (n) and short time series length (T) – where conventional ML is inconsistent due to the effects of incidental parameters. The authors appealed to standard regularity conditions for the asymptotic theory of FDMLE, and used Newton–Raphson optimization in simulations to compute the FDMLE. Their simulations confirmed the superior performance of the FDMLE in terms of bias, root mean square error, test accuracy and power over a range of commonly used panel estimators. HPT do note that FDMLE "sometimes breaks down completely" giving negative variance estimates and estimates of the autoregressive coefficient greater than unity but they "skipped those replications altogether" and provided no analysis of these anomalies.

The present work will explain these anomalies and make it clear why standard asymptotic arguments do not apply to derive the limit theory of the FDMLE. The most closely related work to the present paper is Kruiniger (2008). Kruiniger derived asymptotics for the FDMLE in the panel AR(1) model with large nT (i.e., for n or T large or both n and T large) for the stationary case, and with large n and arbitrary T for the unit root case. Though first differencing uses up one observation for each panel, there appears to be no serious information loss in comparison with other methods like ML because one degree of freedom is needed in conventional ML to identify each individual intercept. Curiously, the asymptotics that are now available speak to the opposite, although this has not so far been discussed in the literature. Indeed, for AR(1) panels with large n, large T and a unit root, the LSDV estimator (which is the MLE under normality of the idiosyncratic error, conditional on initial observations and without any restriction of covariance stationarity) is known to have a $N(0, \frac{51}{5})$ limit distribution when the bias is corrected (Hahn and Kuersteiner, 2002). By contrast, the FDMLE is also asymptotically normal, has no asymptotic bias and its limit variance is 8 (Kruiniger, 2008), thereby producing an asymptotic gain in efficiency at unity over bias corrected LSDV. This reduction in asymptotic variance between the two ML approaches is partly explained by the fact that the FDMLE uses a stationarity condition for the differenced data in setting up the likelihood. Such a condition does not allow for the fact that differenced data is explosive when the AR coefficient exceeds unity, thereby leading to an implied restriction on the model and parameter space that affects both finite sample and asymptotic behavior.

Recent work by Han et al. (2011, forthcoming) shows that there are other estimators involving difference transformations that have performance superior to the bias corrected MLE in dynamic panels. These authors give a panel fully aggregated estimator (FAE) that aggregates the effects of a full set of differences in a simple linear regression framework. The panel FAE has a limiting $N(0, 9)$ distribution after centering and standardization, and like the FDMLE is more efficient asymptotically than the bias corrected MLE with no stationarity restriction imposed (i.e., the bias corrected LSDV) for the autoregressive coefficient in a vicinity of unity. There is much other recent work on dynamic panel models, but none of that work relates to the issues connected with the FDMLE procedure that are discussed in the present paper.

For all the attractive properties of FDMLE, some of its most important features have not been noted or studied in the literature. These features, as we demonstrate here, play a critical role in the asymptotic theory and in the finite sample performance of the estimator. First and most importantly, the 'likelihood' function considered in the panel literature that is used for numerical computation of the FDMLE is *not* in fact the correct likelihood function over the whole domain. As indicated above, it is a pseudo-likelihood based on extending the stationary likelihood outside its natural domain of definition to a bounded part of the nonstationary region. Second, this pseudo 'likelihood' function can behave so wildly that numerical maximization procedures can often fail to identify the global maximum. These two issues combine to make a careful analytical treatment of FDMLE very difficult. On the one hand, the asymptotic theory depends subtly on the (rapidly changing) form of the likelihood function near its natural upper boundary which arises from the extension of the stationary likelihood. On the other hand, the wild behavior of the likelihood itself often compromises the numerical evaluation of the FDMLE, giving rise to anomalous results such as those reported above.

The present paper explains these pathologies and their material impact on the finite sample distribution and limit distribution of the FDMLE. We also show how the effects of this anomaly diminish when the FDML is applied to dynamic panel data as the cross-sectional dimension increases.

The next section lays out the model, notation and discusses the FDML 'likelihood'. Section 3 examines the anomaly that arises when the data are persistent, considering in turn the time series ($n = 1$), panel ($n > 1$) and panel asymptotic case ($n \to \infty$). Section 4 concludes. Proofs are given in the Appendix and reference is made to the original version of this paper (Han and Phillips, 2010) for further technical details. Throughout the remainder of the paper it will be convenient to use the notation $T_m = T - m$ and $\tilde{T}_m = T + m$.

2. Model, notation and the FDMLE

We consider a Gaussian panel y_{it} generated by the simple panel dynamic model $y_{it} = \eta_i(1 - \rho_0) + \rho_0 y_{it-1} + \varepsilon_{it}$, where $\varepsilon_{it} \sim iid\ N(0, \sigma_0^2)$ and $-1 < \rho_0 \leq 1$.[1] Suppose that y_{it} is observed for $i = 1, \ldots, n$ and $t = 0, \ldots, T$.

The likelihood function is derived from the joint distribution of $\Delta y_i := (\Delta y_{i1}, \ldots, \Delta y_{iT})'$. Under the stationarity assumption for Δy_{it}, we have

$$\Delta y_i \sim N\left(0, \sigma_0^2 C_T(\rho_0)\right), \tag{1}$$

where $C_T(\rho_0)$ is a Toeplitz matrix whose leading row is formed from the elements $\frac{1}{1+\rho_0}\{2, -(1-\rho_0), -\rho_0(1-\rho_0), \ldots, -\rho_0^{T-2}(1-\rho_0)\}$. Direct evaluation leads to $\det C_T(\rho_0) = J_T(\rho_0)/(1 + \rho_0)$, where $J_T(\rho) = \tilde{T}_1 - T_1\rho$ (e.g., Galbraith and Galbraith, 1974; HPT, 2002; Kruiniger, 2008; Han, 2007). Thus, for $-1 < \rho \leq 1$ and $\sigma^2 > 0$, the log-likelihood function for Δy_i is

$$\ln L(\rho, \sigma^2) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 - \frac{n}{2}\ln\left[\frac{J_T(\rho)}{1+\rho}\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\Delta y_i' C_T(\rho)^{-1}\Delta y_i. \tag{2}$$

[1] The analysis can be extended to the model where y_{it} is replaced with $y_{it} - \beta' x_{it}$ and x_{it} contains exogenous regressors. The focus in the present paper is on the estimation of ρ and the peculiarities of its limit theory. Asymptotics for the corresponding estimates of β may be derived in a standard way and are not discussed here.