

Incidental Parameters and Dynamic Panel Modeling*

Hyungsik Roger Moon Benoit Perron
University of Southern California University of Montreal
& Yonsei University

Peter C.B. Phillips
Yale University, University of Auckland,
University of Southampton & Singapore Management University

November 28, 2013

Abstract

The present chapter considers the role of incidental parameters in capturing idiosyncratic features of individual entities within a panel and their influence on econometric estimation and inference. While incidental parameters relate to a single cross section observation, their presence affects maximum likelihood estimates of other parameters of interest and can even result in inconsistency. The present chapter discusses incidental parameter problems in the context of a prototype dynamic panel regression model under different assumptions on the size of the panel. We explore the effects on estimation and inference regarding the common panel regression coefficients of different specifications of the incidental parameters for the individual effects and time effects.

1 Introduction

Panel data offers great opportunities for empirical research throughout the social and business sciences, as such data has done for many decades in longitudinal medical studies. In economics and business, just as in medicine, there is often intense interest in the effects of policy measures or treatments on individual consumer and firm behavior over time. The pooling of data records across wide panels of individuals has the potential to deliver substantial econometric power

*The authors thank the referee and Editor for helpful comments on the original version of this chapter. Moon acknowledges financial support from FDA of USC, the research award of Lusk Center of Real Estate, and the research award of Yonsei University. Perron acknowledges financial support from SSHRC and FQRSC-ANR. Phillips acknowledges NSF support under Grant Nos. SES 09-56687 and 12-58258.

in estimation through cross section averaging to sharpen estimates of common response patterns.

With these great opportunities for studying individual behavior come many challenges. The present chapter focuses on one of these challenges – the role and effects of incidental parameters that capture the idiosyncratic features of individual entities within a panel. Adding a new individual to a panel brings new idiosyncratic elements to be explained in the data just as it also brings observations that enhance the power of cross section averaging for the common elements of behavior. Exploring the effects of such additions is the subject of this chapter.

The problem of incidental parameters in statistical inference was first pointed out in a classic article by Neyman and Scott (1948)¹. According to their characterization, an incidental parameter only figures in a finite dimensional probability law – thereby involving only a finite number of observations and, in consequence, rendering the corresponding maximum likelihood estimator inconsistent. Interestingly, in its attempt to deliver the best possible estimates of all of the parameters in a model, including incidental parameters that are specific to a single cross section observation, maximum likelihood estimates of other parameters in the model are inevitably affected and may be inconsistent.

In the context of panel data, the incidental parameter problem typically arises from the presence of individual-specific parameters. These may relate to individual consumer, firm, or country fixed intercept (or mean) effects in a panel. They may also involve incidental trends that are specific to each individual in the sample. The challenges presented by incidental parameters are particularly acute in dynamic panels where behavioral effects over time are being measured in conjunction with individual effects. Nickell (1981) discovered that maximum likelihood estimation of such models leads to inconsistent estimates of the parameters that govern the dynamics. The problems are even more acute in nonstationary panels and panels with incidental trends. Much methodological research on dynamic panel modeling has been devoted to understanding these problems and to developing econometric techniques that address them.

A secondary but no less important issue arises in matters of dynamic model specification. Curiously, some standard model selection procedures such as the BIC information criterion are inconsistent in the presence of incidental parameters. The problem was first identified in a simple example by Stone (1979) and later shown to apply in all dynamic panel models with fixed effects. Model choice is a fundamental aspect of good empirical model building and much applied econometric work relies heavily on standard selection methods such as BIC. The failure of these methods in the context of dynamic panels therefore is a major obstacle to good empirical practice. Approaches to overcome the inconsistency of BIC and related Bayesian procedures have been developed only very recently. Typically one needs to use a special prior in the conventional approach (see Berger et al, 2003). Or in a Kullback Leibler approach, one can increase the penalty. Lee (2012) has some discussion of these possibilities in econometrics.

¹See Lancaster (2000) for an historical overview.

A wider literature is now emerging - see Casella and Moreno (2006, 2009) and Moreno, Girón and Casella (2010). Recent work by Han, Phillips and Sul (2013) has shown that BIC is also inconsistent as an order estimator in dynamic panels even when no fixed effects are present. So the issues presented by model choice in dynamic panels are wider than those involving incidental parameters alone.

The present chapter will discuss these issues in the context of a prototype model under different assumptions on the size of the panel and the properties of the variables in the panel. The prototype of central focus is a simple linear dynamic panel regression with various forms of unobserved characteristics. The model includes only a lagged dependent variable as a regressor but it is the kernel of all dynamic panel models. As it stands, the prototype model is too simplistic for empirical applications where it is often necessary to control for the effects of additional regressors. But this simple case helps us to understand and illustrate the nature of incidental parameter problems in dynamic panel regressions under various fixed effect assumptions. Some of the original papers that are cited in each section analyze an extended framework with extra regressors, and readers are encouraged to refer to these as needed.

The prototype model is as follows. Suppose a double indexed random variable y_{it} is observed over N cross section individuals $i = 1, \dots, N$ and T time periods $t = 1, \dots, T$. We consider two forms of dynamic specification. The first is the component model where the observed data comprise an individual and time effect plus a disturbance that follows an autoregressive model:

$$y_{it} = \delta(\lambda_i, f_t) + \varepsilon_{it} \quad (1)$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it}. \quad (2)$$

The second is an augmented regression model form:

$$y_{it} = \rho y_{it-1} + \alpha(\lambda_i, f_t) + u_{it}, \quad (3)$$

These two formulations are related when $\alpha(\lambda_i, f_t) = \delta(\lambda_i, f_t) - \rho\delta(\lambda_i, f_{t-1})$ and the error dynamics in (2) are absorbed into (1) giving (3). The distinction is particularly relevant when the data are nonstationary.

In models (1) and (3), λ_i and f_t represent individual specific effects and time specific effects, respectively. In this chapter, we treat λ_i and f_t as fixed, so that they can be arbitrarily correlated with the initial condition and are unknown parameters that need to be estimated. While these individual and time specific parameters are relevant and important, it is usually the parameter ρ that governs the behavioral dynamics that is of primary interest in practical work. This chapter therefore focuses on issues of statistical estimation and inference concerning the common behavioral parameter ρ .

2 Formulations and Problems of Interest

Several forms of specification for fixed effects have been used in practical work. In the augmented regression model (3), the most common form is a simple additive individual effect with $\alpha(\lambda_i, f_t) = \lambda_i$. This specification is typically used

to model unobservable differences in characteristics that do not vary over time but that would invalidate results if they were not taken into account in the regression. For example, intrinsic ability in a wage equation is an unobserved individual characteristic that materially affects the wage of an individual. Inclusion of such individual effects in regression helps to model situations where individuals with the same observable characteristics have different outcomes (here, wages) for reasons that the econometrician cannot observe. In such cases, the individual effect serves the role of an individual specific dummy variable in the regression.

Other common specifications of fixed effects in the augmented regression are combined individual-specific and time-specific effects, as in the additive form $\alpha(\lambda_i, f_t) = \lambda_i + f_t$, and interactive fixed effects, as in the multiplicative form $\alpha(\lambda_i, f_t) = \lambda_i f_t$ (or $\lambda_i' f_t$ if λ_i and f_t are vectors). The time-specific variable f_t may represent nonstationarity (or time evolution) in the time series of the panel y_{it} , or it may represent a common shock that provides a source of cross-sectional dependence in the panel y_{it} . For example, all individual consumers in a panel may face a common interest rate shock or individual firms may all be subjected to a common exchange rate shock.

It is also natural to allow for the possibility of time trends in the variables of a panel that might be captured through a deterministic trend. Such effects can be incorporated in a component model of the form $\delta(\lambda_i, f_t) = \lambda_{i0} + \lambda_{i1}t$, so that each unit in the panel embodies a linear trend with its own unit-specific slope. This formulation is called an incidental trend effect. Such effects commonly appear in panel models of cross country economic growth.

The impact of incidental parameters on dynamic panel estimation was studied by Nickell (1981) who considered a panel autoregression of order one with common coefficient ρ in the presence of individual-specific intercepts $\alpha(\lambda_i, f_t) = \lambda_i$. Analytical expressions showed that the pooled least square (OLS) estimator or Gaussian MLE is asymptotically (as the cross section dimension $N \rightarrow \infty$) biased downward (if $\rho > 0$), and that the bias decreases as the time series sample size T increases. Thus, using OLS or MLE results in inconsistent estimates of ρ when $N \rightarrow \infty$ whenever there is a finite time series sample for each cross-sectional unit. The phenomenon arises as a consequence of having only a finite number of observations from which to estimate the individual-specific parameter λ_i , which contaminates the estimation of ρ . In effect, as intimated in the Introduction, maximum likelihood in attempting to get the best estimates of all the parameters in the model (including the individual effects) fails to achieve consistent estimation even for the common parameter ρ .

The following sections consider several instances of incidental parameter problems in dynamic panels, as well as various solutions that have been proposed to circumvent them. The remainder of the chapter is divided into four parts. The next section deals with the estimation of the panel autoregressive coefficient ρ , while section 4 is concerned with inference about ρ and tests of specific null hypotheses such as $H_0 : \rho = \rho_0$, i.e., that ρ equals a certain constant value of interest such as zero or unity. Much attention has been given to the unit root case where the hypothesis of interest is that members of the panel follow

dynamic paths described by random walks in which $\rho = 1$. We discuss recent work on nonlinear panels and on model selection in dynamic panels in Sections 5 and 6. Throughout the chapter we consider large N asymptotics and cases where T is finite and $T \rightarrow \infty$. Only Classical methods are discussed, although Bayesian principles underlie some of the model selection criteria.²

3 Estimating ρ with Incidental Parameters

3.1 Individual Fixed Effects: $\alpha(\lambda_i, f_t) = \lambda_i$

We consider the dynamic panel regression model (3) with time invariant additively separable fixed effects, $\alpha(\lambda_i, f_t) = \lambda_i$ and $|\rho| < 1$:

$$y_{it} = \rho y_{it-1} + \lambda_i + u_{it}, \quad (4)$$

where y_{i0} is the initial condition and the error term u_{it} is uncorrelated with the $\{y_{it-s} : s \geq 1\}$. The individual effect λ_i in model (4) is fixed in the sense that it can be arbitrarily correlated with the initial condition y_{i0} and we treat these individual effects as parameters to be estimated. The parameter of interest is the common dynamic coefficient ρ .

The Gaussian quasi-maximum likelihood estimator (QMLE) of ρ is the maximum likelihood estimator when $u_{it} \sim iidN(0, \sigma^2)$ and the λ_i are nuisance parameters. Nickell (1981) showed that the QMLE of ρ is inconsistent as $N \rightarrow \infty$ with T fixed and finite, due to the presence of the incidental parameters λ_i in the regression (4).

3.1.1 Explaining the Nickell Bias

To explain the source of the asymptotic bias, it is useful to simplify the algebra before developing the large N asymptotics in detail. The Gaussian QMLE of ρ is equivalent to pooled least squares on the panel and has the form

$$\hat{\rho} = \left(\sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})^2 \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})(y_{it} - \bar{y}_i), \quad (5)$$

where $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$ and $\bar{y}_{i,-1} = T^{-1} \sum_{t=1}^T y_{it-1}$. In view of the demeaning transformation within the panel for each individual i , the estimator is also called the within estimator (or fixed effect or least squares dummy variable estimator) in the literature. In taking the time series mean from the data, the within transformation eliminates the fixed effects λ_i because

$$y_{it} - \bar{y}_i = \rho(y_{it-1} - \bar{y}_{i,-1}) + u_{it} - \bar{u}_i. \quad (6)$$

²For Bayesian analysis of dynamic panels, one could refer to Lancaster (2002) for example.

When $|\rho| < 1$, the solution of y_{it} in (4) is

$$y_{it} = \frac{\lambda_i}{1 - \rho} + y_{it}^0, \text{ where } y_{it}^0 = \rho y_{it-1}^0 + u_{it} = \sum_{j=0}^{\infty} \rho^j u_{it-j}.$$

It follows that $y_{it} - \bar{y}_i = y_{it}^0 - \bar{y}_i^0$ and $y_{it-1} - \bar{y}_{i,-1} = y_{it-1}^0 - \bar{y}_{i,-1}^0$. The estimation error in $\hat{\rho}$ therefore has the following form

$$\begin{aligned} \hat{\rho} - \rho &= \left(\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0)^2 \right)^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0) (u_{it} - \bar{u}_i) \\ &= \left(\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0)^2 \right)^{-1} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0) u_{it}. \end{aligned} \quad (7)$$

Since $\sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0)^2$ is *iid* over i , the ergodic theorem implies that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0)^2 &\rightarrow_{a.s.} \mathbb{E} \left\{ \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0)^2 \right\} = T \mathbb{E} (y_{it-1}^0 - \bar{y}_{i,-1}^0)^2 \\ &= T \mathbb{E} (y_{it-1}^0)^2 - T \mathbb{E} (\bar{y}_{i,-1}^0)^2 = T \frac{\sigma^2}{1 - \rho^2} - \frac{1}{T} \sum_{t,s=1}^T \mathbb{E} (y_{it} y_{is}) \\ &= T \frac{\sigma^2}{1 - \rho^2} - \frac{1}{T} \sum_{t,s=1}^T \rho^{|t-s|} \frac{\sigma^2}{1 - \rho^2} \\ &= T \frac{\sigma^2}{1 - \rho^2} - \frac{1}{T} \sum_{j=-T+1}^{T-1} \rho^{|j|} \left[\sum_{t,s=1}^T \mathbf{1}\{t-s=j\} \right] \frac{\sigma^2}{1 - \rho^2} \\ &= T \frac{\sigma^2}{1 - \rho^2} - \frac{1}{T} \sum_{j=-T+1}^{T-1} \rho^{|j|} [T - |j|] \frac{\sigma^2}{1 - \rho^2} \\ &= T \frac{\sigma^2}{1 - \rho^2} + O(1), \text{ because } |\rho| < 1 \end{aligned} \quad (8)$$

and

$$\begin{aligned}
\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0) u_{it} &\rightarrow_{a.s.} \mathbb{E} \left\{ \sum_{t=1}^T (y_{it-1}^0 - \bar{y}_{i,-1}^0) u_{it} \right\} = \sum_{t=1}^T \mathbb{E} (y_{it-1}^0 u_{it} - \bar{y}_{i,-1}^0 u_{it}) \\
&= - \sum_{t=1}^T \mathbb{E} (\bar{y}_{i,-1}^0 u_{it}), \text{ because } \mathbb{E} y_{it-1}^0 u_{it} = 0 \\
&= - \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\sum_{s=1}^T y_{is-1}^0 u_{it} \right) = - \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\sum_{s=t+1}^T y_{is-1}^0 u_{it} \right) \\
&= - \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left(\sum_{s=t+1}^T y_{is-1}^0 u_{it} \right) = - \frac{1}{T} \sum_{t=1}^T \mathbb{E} \left\{ \sum_{s=t+1}^T \sum_{j=0}^{\infty} \rho^j u_{is-1-j} u_{it} \right\} \\
&= - \sigma^2 \frac{1}{T} \sum_{t=1}^T \sum_{j=0}^{T-t-1} \rho^j \text{ since } \mathbb{E} \{ u_{is-i-j} u_{it} \} = \sigma^2 \mathbf{1} \{ s = t + 1 + j \} \\
&= - \frac{\sigma^2}{1-\rho} + \frac{\sigma^2}{1-\rho} \frac{1}{T} \sum_{t=1}^T \rho^{T-t} = - \frac{\sigma^2}{1-\rho} + \frac{\sigma^2}{1-\rho} \frac{1}{T} \sum_{j=0}^{T-1} \rho^j \\
&= - \frac{\sigma^2}{1-\rho} + O\left(\frac{1}{T}\right). \tag{9}
\end{aligned}$$

It follows directly from (8) and (9) that as $N \rightarrow \infty$

$$\hat{\rho} - \rho \rightarrow_{a.s.} \frac{-\frac{\sigma^2}{1-\rho} + O\left(\frac{1}{T}\right)}{T \frac{\sigma^2}{1-\rho^2} + O(1)} = - \frac{1+\rho}{T} + O\left(\frac{1}{T^2}\right), \tag{10}$$

and $\hat{\rho}$ is inconsistent. The inconsistency is non-trivial for small T . For example when $\rho = 0.5$ and $T = 5$, the asymptotic bias is $-\frac{1+\rho}{T} = -0.3$, so in this case $\hat{\rho} \rightarrow_{a.s.} 0.2$. When N is large, because of the small variance in $\hat{\rho}$, almost all of the distribution of $\hat{\rho}$ then lies to the left of the true value $\rho = 0.5$ and confidence intervals have a coverage probability close to zero. In short wide panels (with T small and N large), the inconsistency can therefore have dramatic effects on inference.

The bias expression (10) holds only for the stationary panel case where $|\rho| < 1$. The unit root case $\rho = 1$ is handled in a similar way and we obtain (see Phillips and Sul, 2007)

$$\hat{\rho} - \rho \rightarrow_{a.s.} - \frac{3}{T} + O\left(\frac{1}{T^2}\right), \tag{11}$$

so the bias effects are magnified in the unit root case and there is no continuity in the asymptotic bias expression as ρ moves to unity. The bias in the unit case can clearly be very large when T is small. For example, when $T = 3$ and 4 the respective bias is -1 and -0.75 , almost sufficient to change the sign of $\hat{\rho}$.

As is apparent from the calculation leading to (9), u_{it} and $\bar{y}_{i,-1}^0$ are correlated and it is this correlation that produces the inconsistency in $\hat{\rho}$ as $N \rightarrow \infty$ for fixed T . Thus, it is the removal of the fixed effects by the within regression transformation of demeaning (manifest in the presence of $\bar{y}_{i,-1}^0$ in (7)) that induces the correlation. Notice also that the QMLE of λ_i is $\hat{\lambda}_i = \bar{y}_i - \hat{\rho}\bar{y}_{i,-1}$, so that there is a further induced bias in the estimation of the individual effects. The numerator of $\hat{\rho} - \rho$, $\sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1}) u_{it}$, is the score function of the concentrated Gaussian quasi-likelihood function and it too is biased with consequential effects on testing. Thus, the incidental parameter problem arising from the presence of fixed effects in (4) has a manifold deleterious impact on estimation and inference in dynamic panel regression.

3.1.2 Panels with Fixed T : First Differencing and Instrumental Variables

When the number of time series observations T is fixed, methods to overcome the incidental parameter problem described in the previous analysis are well studied and many solutions have been suggested since the early 1980's. Excellent overviews of the subject are available, including the chapter by Arellano and Honore (2001), and the books by Arellano (2003), Hsiao (2003), and Baltagi (2008).

The most common approach is to rely on first differencing to eliminate the fixed effect instead of the within transformation. The resulting equation is then:

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}. \quad (12)$$

However, while this transformation removes the individual effect, it introduces a moving average component of order 1 in the error term which brings about correlation with the regressor in (12). However, as noted by Anderson and Hsiao (1982) past values of the dependent variable satisfy the necessary moment conditions for a valid instrument, i.e.

$$E(\Delta u_{it} y_{it-s}) = E((u_{it} - u_{it-1}) y_{it-s}) = 0, \quad \text{for any } s > 1. \quad (13)$$

They proposed instrumental variable estimation with y_{t-2} as instrument, leading to

$$\hat{\rho}_{IV} = \left(\sum_{i=1}^N \sum_{t=1}^T y_{it-2} \Delta y_{it-1} \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T y_{it-2} y_{it}$$

which is consistent and asymptotically normal as $N \rightarrow \infty$ for T fixed.

Of course, because all lagged values of y_{it-1} are valid instruments, one could use many more instruments among the overall $T(T-1)/2$ moment conditions in the (13) class. These may be mobilized in a GMM framework as suggested by Arellano and Bond (1991). Ahn and Schmidt (1995) further added the $T-2$ moments $E(\Delta u_{it} u_{iT}) = 0$ since u_{it} is assumed to be serially uncorrelated. Ahn and Schmidt also show that the estimator that uses these $T(T-1)/2 + (T-2)$ moments uses all the moments implied by the basic assumptions, and that

the resulting estimator correspondingly reaches the semi-parametric efficiency bound of Chamberlain (1982,1984).

Han, Phillips, and Sul (2012a)³ generalize this idea and introduce the new concept of X-differencing to generate moment conditions. The procedure eliminates the fixed effects like conventional first differencing while making the regressor and error uncorrelated after the transformation. Hence, there is no need for instrumental variables, and the method does not suffer from the weak identification problem that arises as the autoregressive parameter approaches unity (a problem originally noted by Blundell and Bond, 1998). The method combines the basic equation (4) with the forward-looking regression

$$y_{is} = \lambda_i + \rho y_{is+1} + \varepsilon_{is}^* \quad (14)$$

where $\varepsilon_{is}^* = \varepsilon_{is} - \rho(y_{is+1} - y_{is-1})$, which is uncorrelated with y_{is+1} if ε_{it} is serially uncorrelated and uncorrelated with the individual effect $E(\lambda_i \varepsilon_{is}) = 0$, though this condition is not needed for the properties of the estimator as the individual effects are eventually eliminated. The same orthogonality condition applies when replacing $s + 1$ by $t > s$.

Subtracting (14) from (4), leads to the simple regression equation

$$y_{it} - y_{is} = \rho(y_{it-1} - y_{is+1}) + (\varepsilon_{it} - \varepsilon_{is}^*) \quad (15)$$

where the regressor and error are uncorrelated for any $s < t - 1$ and any $-1 < \rho \leq 1$ so that the approach accommodates the unit root case $\rho = 1$ within the same framework. The X-differencing terminology is suggested by virtue of the fact that the regressand is differenced by $X = t - s$ periods while the regressor is differenced by $X - 2$ periods. All admissible values of $s = 1, \dots, t - 3$ or $X = 3, \dots, t - 1$ can be considered.

Based on these X-differences, Han, Phillips and Sul construct a panel fully aggregated estimator (PFAE) as the pooled regression estimator in (15) for all i , t , and s

$$\hat{\rho}_{PFAE} = \left(\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-3} (y_{it-1} - y_{is+1})^2 \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^{t-3} (y_{it-1} - y_{is+1}) (y_{it} - y_{is}).$$

This estimator is consistent and asymptotically normal as long as the number of observations NT goes to infinity. Thus, it is consistent for fixed T as $N \rightarrow \infty$. More importantly, given the orthogonality between the error and regressor after the transformation, it is essentially unbiased in finite samples as confirmed in simulations. The FAE estimator has other appealing properties, including improved efficiency when ρ is in the vicinity of unity.

Other transformations have been considered to eliminate the fixed effects and allow for consistent estimation of ρ even with a finite number of time series observations. For example, Arellano and Bover (1995) proposed forward

³See also Han and Phillips (2010) and Han Phillips and Sul (2011).

orthogonal differences:

$$y_{it}^* = \sqrt{\frac{T-t}{T-t+1}} \left[y_{it} - \frac{1}{T-t} (y_{it+1} + \dots + y_{iT}) \right]. \quad (16)$$

For detailed results of the analysis of the various IV and GMM estimators of ρ when T is fixed, readers may refer to several standard textbook treatments such as Arellano (2003), Baltagi (1995), and Hsiao (2003). Another approach using indirect inference methods has been suggested recently by Gourieroux, Phillips and Yu (2010).

3.1.3 When T is Large

When the time series dimension of the panel is large, Hahn and Kuersteiner (2002) proposed to use joint $(N, T) \rightarrow \infty$ asymptotics to characterize the bias of the fixed effect estimator that arises from the incidental parameters λ_i . More specifically, they derived the limit distribution of $\hat{\rho}$ by allowing $N, T \rightarrow \infty$ jointly under the rate condition $\frac{N}{T} \rightarrow \kappa^2$, where $0 < \kappa < \infty$, so that N and T pass to infinity at the same rate. For expositional simplicity here⁴, we assume that $u_{it} \sim iid(0, \sigma_u^2)$ across i and over t with finite fourth moments. We further assume that $\frac{1}{N} \sum_{i=1}^N y_{i0}^2 = O_p(1)$ and the fixed effects λ_i satisfy $\frac{1}{N} \sum_{i=1}^N \lambda_i^2 = O_p(1)$.

The centered and normalized within estimator (5) has the form

$$\sqrt{NT}(\hat{\rho} - \rho) = \left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})^2 \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1}) u_{it}. \quad (17)$$

As $N, T \rightarrow \infty$, the denominator of $\sqrt{NT}(\hat{\rho} - \rho)$ has the following limit

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})^2 \rightarrow_p \frac{\sigma_u^2}{1 - \rho^2}, \quad (18)$$

mirroring our earlier result (8) for fixed T . Defining $u_{it}(\rho) = \sum_{s=0}^{t-1} \rho^s u_{it-s}$, the numerator of (17) decomposes as follows

$$\begin{aligned} & \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1}) u_{it} \\ &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T u_{it-1}(\rho) u_{it} - \sqrt{\frac{N}{T}} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T u_{it-1}(\rho) u_{is} \right) + o_p(1). \end{aligned}$$

Since $u_{it-1}(\rho) u_{it}$ is a martingale difference, the first term satisfies an extended version of the martingale central limit theorem (c.f. Phillips and Moon, 1999)

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T u_{it-1}(\rho) u_{it} \Rightarrow \mathcal{N} \left(0, \frac{\sigma_u^4}{1 - \rho^2} \right),$$

⁴Hahn and Kuersteiner (2002) also considered the more general case where the regression errors may be conditionally heteroskedastic.

whereas the second term converges in probability to a constant

$$\sqrt{\frac{N}{T}} \left(\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T u_{it-1}(\rho) u_{is} \right) \rightarrow_p \kappa \frac{\sigma_u^2}{1-\rho},$$

which mirrors our earlier result (9) for the numerator as $N \rightarrow \infty$ for fixed T . Therefore, as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$, the weak limit of the numerator of $\sqrt{NT}(\hat{\rho} - \rho)$ is

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1})(u_{it} - \bar{u}_i) \Rightarrow \mathcal{N} \left(-\kappa \frac{\sigma_u^2}{1-\rho}, \frac{\sigma_u^4}{1-\rho^2} \right). \quad (19)$$

Combining (18) and (19) as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$ and where $0 < \kappa < \infty$, we have

$$\sqrt{NT}(\hat{\rho} - \rho) \Rightarrow \mathcal{N}(-\kappa(1+\rho), 1-\rho^2). \quad (20)$$

An important aspect of the limit distribution of the fixed estimator is the bias involved in the miscentred normal limit of (20). This bias, like that for the $N \rightarrow \infty$ case, is due to the presence of the incidental parameters λ_i in (4) and arises from the limiting correlation apparent in (19) that is induced by the (within) transformation to eliminate the nuisance parameters. A finite sample implication of (20) is that the bias of $\hat{\rho}$ can be approximated by $-\frac{1+\rho}{T}$, just as suggested in the earlier limit (10) for fixed T .

Accordingly Hahn and Kuersteiner (2002) suggested the following bias-corrected estimator

$$\check{\rho} = \hat{\rho} + \frac{1}{T}(1 + \hat{\rho}) = \frac{T+1}{T}\hat{\rho} + \frac{1}{T},$$

which, as $\frac{N}{T} + \frac{1}{N} \rightarrow \kappa^2$ with $0 < \kappa < \infty$, has the following correctly centred limit distribution

$$\sqrt{NT}(\check{\rho} - \rho) \Rightarrow \mathcal{N}(0, 1 - \rho^2). \quad (21)$$

This bias corrected estimator is shown in Hahn and Kuersteiner to be asymptotically efficient under Gaussian errors provided $|\rho| < 1$. The unit case $\rho = 1$ is more complex and the limit theory (21) no longer holds. While the limit distribution is still normal, the rate of approach to the limit theory is no longer $O(\sqrt{NT})$ but is instead $O(\sqrt{NT^2})$, reflecting the stronger time series signal in the regressors in the unit root case. Hahn and Kuersteiner proved that $\hat{\rho}$ has the following limit distribution in this case when $\rho = 1$ and $(N, T) \rightarrow \infty$

$$\sqrt{NT^2} \left(\hat{\rho} - \rho + \frac{3}{T} \right) \Rightarrow \mathcal{N} \left(0, \frac{51}{5} \right).$$

Alvarez and Arellano (2003) studied the GMM estimator $\hat{\rho}_{GMM}$ based on the transformed data (16) and using the lagged dependent variables $z_{it} =$

$(y_{it-1}, \dots, y_{i0})$ as IVs expressed as

$$\begin{aligned} \hat{\rho}_{GMM} &= \left(\sum_{t=1}^T \left(\sum_{i=1}^N x_{it}^* z'_{it} \right) \left(\sum_{i=1}^N z_{it} z'_{it} \right)^{-1} \left(\sum_{i=1}^N z_{it} x_{it}^* \right) \right)^{-1} \\ &\quad \times \left(\sum_{t=1}^T \left(\sum_{i=1}^N x_{it}^* z'_{it} \right) \left(\sum_{i=1}^N z_{it} z'_{it} \right)^{-1} \left(\sum_{i=1}^N z_{it} y_{it}^* \right) \right), \end{aligned}$$

where $x_{it}^* = \sqrt{\frac{T-t}{T-t+1}} \left[y_{it-1} - \frac{1}{T-t} (y_{it} + \dots + y_{iT-1}) \right]$. They showed that as $T \rightarrow \infty$ the bias of $\hat{\rho}_{GMM}$ may be approximated by

$$-\frac{1}{N} (1 + \rho)$$

and its limiting distribution is

$$\sqrt{NT} \left(\hat{\rho}_{GMM} - \rho + \frac{1}{N} (1 + \rho) \right) \Rightarrow \mathcal{N} (0, 1 - \rho^2)$$

which can be written as

$$\sqrt{NT} (\hat{\rho}_{GMM} - \rho) \Rightarrow \mathcal{N} (-\kappa^{-1} (1 + \rho), 1 - \rho^2).$$

General Motivation of the Alternative Asymptotics: The fixed effect estimator $\hat{\rho}$ corresponds to the maximum likelihood estimator based on the conditional Gaussian likelihood. To see this, suppose that conditional on y_{i0} and λ_i , $u_{it} \sim iid N(0, \sigma_u^2)$ with known σ_u^2 . Then, the conditional log-likelihood of (y_{iT}, \dots, y_{i1}) on $(y_{i0}, \lambda_i)_i$ is

$$l_{NT}(\rho, \lambda^N) = \sum_{i=1}^N \ln f(y_{iT}, \dots, y_{i1} | y_{i0}, \rho, \lambda_i) = \sum_{i=1}^N \sum_{t=1}^T l_{it}(\rho, \lambda_i),$$

where

$$l_{it}(\rho, \lambda_i) = -\frac{1}{2\sigma_u^2} (y_{it} - \rho y_{it-1} - \lambda_i)^2.$$

Without loss of generality, assume that $\sigma_u^2 = 1$ for simplicity. Then, for given ρ the MLE of λ_i is

$$\hat{\lambda}_i(\rho) = \arg \max_{\lambda} \sum_{t=1}^T l_{it}(\rho, \lambda) = \bar{y}_i - \rho \bar{y}_{i,-1}.$$

Plugging $\hat{\lambda}_i(\rho)$ into the likelihood, we have the (concentrated) profile likelihood of ρ :

$$\begin{aligned} l_{NT}(\rho) &= l_{NT}(\rho, \hat{\lambda}_i(\rho)) = \sum_{i=1}^N \sum_{t=1}^T l_{it}(\rho, \hat{\lambda}_i(\rho)) \\ &= -\frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i - \rho (y_{it-1} - \bar{y}_{i,-1}))^2. \end{aligned}$$

The fixed effect estimator $\hat{\rho}$ is simply the MLE in this case since

$$\hat{\rho} = \arg \max_{\rho} l_{NT}(\rho).$$

Define

$$L_T(\rho) = \lim_N \frac{1}{N} E(l_{NT}(\rho)) = \lim_N \frac{1}{N} \sum_{i=1}^N E \left[\sum_{t=1}^T l_{it}(\rho, \hat{\lambda}_i(\rho)) \right].$$

As follows by standard asymptotic theory for extremum estimators, when $N \rightarrow \infty$ but T is fixed, the MLE $\hat{\rho}$ converges in probability to

$$\hat{\rho} \rightarrow_p \rho_T = \arg \max_{\rho} L_T(\rho).$$

Denote by (ρ_0, λ_i^0) the true parameters of (ρ, λ_i) . Then, since

$$\sum_{t=1}^T l_{it}(\rho, \hat{\lambda}_i(\rho)) = -\frac{1}{2} \sum_{t=1}^T \{(u_{it} - \bar{u}_i) - (\rho - \rho_0)(y_{it-1} - \bar{y}_{i,-1})\}^2$$

we have

$$\begin{aligned} L_T(\rho) &= -\frac{1}{2} (\rho - \rho_0)^2 \lim_N \frac{1}{N} \sum_{i=1}^N E \left(\sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1}) \right)^2 \\ &\quad + (\rho - \rho_0) \lim_N \frac{1}{N} \sum_{i=1}^N E \left(\sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1}) (u_{it} - \bar{u}_i) \right) \\ &\quad - \frac{1}{2} \lim_N \frac{1}{N} \sum_{i=1}^N E \left(\sum_{t=1}^T (u_{it} - \bar{u}_i)^2 \right). \end{aligned} \quad (22)$$

For fixed T , the expected score of the profile likelihood $E \left(\sum_{t=1}^T (y_{it-1} - \bar{y}_{i,-1}) (u_{it} - \bar{u}_i) \right) \neq 0$, and we have

$$\rho_T \neq \rho_0.$$

However, when $T \rightarrow \infty$ as well as $N \rightarrow \infty$, we have

$$\rho_T \rightarrow \rho_0.$$

Arellano and Hahn (2006) observed that when the profile likelihood function $l_{NT}(\rho)$ is smooth, we usually have the expansion

$$\rho_T = \rho_0 + \frac{B}{T} + O\left(\frac{1}{T^2}\right), \quad (23)$$

and the re-centred limit theory

$$\sqrt{NT}(\hat{\rho} - \rho_T) \Rightarrow \mathcal{N}(0, \Omega)$$

for some $\Omega > 0$. Then, as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$, where $0 < \kappa < \infty$, we have

$$\begin{aligned}\sqrt{NT}(\hat{\rho} - \rho_0) &= \sqrt{NT}(\hat{\rho} - \rho_T) + \sqrt{NT}\frac{B}{T} + O\left(\sqrt{\frac{N}{T^3}}\right) \\ &\Rightarrow \mathcal{N}(\kappa B, \Omega).\end{aligned}$$

Here the bias B of the fixed effect estimator is characterized as the bias of the asymptotic distribution under these joint asymptotics.

3.1.4 Alternative Bias Correction Method

Dhaene and Jochmans (2012) proposed a jackknife method to reduce the order of the bias of the fixed effect estimator in nonlinear dynamic panel regression models. To apply their ideas to the dynamic linear set up in (4), recall that the pseudo true value ρ_T that maximizes the limit of the profile likelihood $L_T(\rho)$ in (22) is $\rho_T = \rho_0 + \frac{B}{T} + O\left(\frac{1}{T^2}\right)$. We now split the $(N \times T)$ panel into two $(N \times \frac{T}{2})$ dimensional pieces and denote by $\hat{\rho}_1$ and $\hat{\rho}_2$ the fixed effect estimators in these respective subpanels. Define

$$\bar{\rho}_{1/2} = \frac{1}{2}(\hat{\rho}_1 + \hat{\rho}_2).$$

Dhaene and Jochmans's bias corrected estimator is then based on the usual jackknife formula

$$\hat{\rho}_{1/2} = 2\hat{\rho} - \bar{\rho}_{1/2}.$$

We can expect this estimator to correct the bias B because by using the expansion (23) we have

$$\begin{aligned}\hat{\rho}_{1/2} &= 2\rho_0 + \frac{2B}{T} + O\left(\frac{1}{T^2}\right) - \frac{1}{2}\left(2\rho_0 + \frac{4B}{T} + O\left(\frac{1}{T^2}\right)\right) \\ &= \rho_0 + O\left(\frac{1}{T^2}\right).\end{aligned}$$

Dhaene and Jochmans (2012) showed that this idea can be applied to a more general nonlinear dynamic panel regression models with fixed effects and show how to reduce the bias to a higher order.

3.2 Additive Individual Effects and Time Effects: $\alpha(\lambda_i, f_t) = \lambda_i + f_t$

Hahn and Moon (2006) extended Hahn and Kuersteiner's (2002) results by considering time effects as additional incidental parameters in the model

$$y_{it} = \rho y_{it-1} + \lambda_i + f_t + u_{it}, \quad (24)$$

where the model satisfies the conditions in the previous section except that it now includes time specific fixed effects f_t . In many empirical applications,

the time effect f_t is included to model a simple form of nonstationarity in the time series y_{it} or to represent an aggregate shock (e.g., a common macro shock) that is common to all the cross-section units. In the latter case, when the common shock f_t is random, the cross sectional observations y_{it} have cross-sectional dependence. In model (24) incidental parameters exist in both the cross sectional direction (λ_i) and the time series direction (f_t).

Define $\bar{y}_{\cdot,t} = \frac{1}{N} \sum_{i=1}^N y_{it}$, $\bar{y}_{i,\cdot} = \frac{1}{T} \sum_{t=1}^T y_{it}$, $\bar{y} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it}$, $\bar{y}_{i,-1} = \frac{1}{T} \sum_{t=1}^T y_{it-1}$, $\bar{y}_{-1} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it-1}$. Similar definitions are used for λ , \bar{f} , $\bar{u}_{\cdot,t}$, $\bar{u}_{i,\cdot}$, and \bar{u} . In this model, the fixed effect estimator is

$$\hat{\rho} = \left(\sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{\cdot,t-1} - \bar{y}_{i,-1} + \bar{y}_{-1})^2 \right)^{-1} \times \left(\sum_{i=1}^N \sum_{t=1}^T (y_{it-1} - \bar{y}_{\cdot,t-1} - \bar{y}_{i,-1} + \bar{y}_{-1}) (y_{it} - \bar{y}_{\cdot,t} - \bar{y}_{i,\cdot} + \bar{y}) \right).$$

Hahn and Moon (2006) showed that under joint asymptotics as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$, where $0 < \kappa < \infty$, $\sqrt{NT}(\hat{\rho} - \rho)$ has the limit distribution

$$\sqrt{NT}(\hat{\rho} - \rho) \Rightarrow \mathcal{N}(-\kappa(1 + \rho), 1 - \rho^2),$$

identical to the limit given earlier in (20). Hence, the asymptotic bias of the fixed effect estimator with $\lambda_i + f_t$ is the same as the asymptotic bias of the fixed effect estimator only with λ_i .

In this case the fixed effect estimator $\hat{\rho}$ eliminates f_t by taking out the cross-sectional mean of the panel data y_{it} . The regressor after this demeaning transformation becomes $y_{it-1} - \bar{y}_{\cdot,t-1}$ and the error becomes $u_{it} - \bar{u}_{\cdot,t}$. From a time series perspective the regressor $y_{it-1} - \bar{y}_{\cdot,t-1}$ is still predetermined and uncorrelated with the error $u_{it} - \bar{u}_{\cdot,t}$. From a cross-sectional perspective, the influence of the time specific effect f_t in the regressor $y_{it-1} - \bar{y}_{\cdot,t-1}$ and the error $u_{it} - \bar{u}_{\cdot,t}$ is removed. As a consequence, the limit of the estimator $\hat{\rho}$ is not affected by the presence of f_t .

This finding confirms that the asymptotic bias of the fixed effect estimator in the linear dynamic panel regression model is sourced in the presence of the individual effect λ_i not the presence of the time effect f_t when these components enter the model in an additive and separable form. However, when the panel model is nonlinear and/or λ_i and f_t enter the model in a more general functional form, it may be the case that both λ_i and f_t contribute to the asymptotic bias of the fixed effect estimator as we will discuss in the following sections.

3.3 Interactive Fixed Effects: $\alpha(\lambda_i, f_t) = \lambda_i' f_t$

In this section, we discuss the case where the fixed effects take a multiplicative form involving λ_i and f_t , viz.,

$$y_{it} = \rho y_{it-1} + \lambda_i' f_t + u_{it}, \quad (25)$$

where λ_i and f_t are unknown fixed effects for i and t , respectively and u_{it} are idiosyncratic shocks. We denote by $\lambda_i^0, f_t^0, \rho_0$ the true values of λ_i, f_t , and ρ , respectively. In this section we assume the dimension of f_t and λ_i are known, say R^0 . Also assume that u_{it} are independent across i and t with mean zero and higher moments finite. The multiplicative form of the fixed effects appearing in $\lambda_i' f_t$ are often called interactive fixed effects or common factors in the literature.

The linear panel regression with interactive fixed effects was studied by Kiefer (1980), Lee (1991), Ahn, Lee, and Schmidt (2001), and Bai (2009) when the regressors are strictly exogenous with respect to u_{it} , and by Holtz-Eakin, Newey and Rosen (1988), Phillips and Sul (2003), and Moon and Weidner (2010, 2013) when the regressors are lagged dependent variables.⁵

3.3.1 Quasi-Differencing Approach

Holtz-Eakin, Newey, and Rosen (1988) suggest that the interactive fixed effects be eliminated by taking a quasi-difference of the data. Suppose that $R_0 = 1$ for expositional convenience and that we normalize $f_1 = 1$. Since

$$\frac{y_{it}}{f_t} = \rho \frac{y_{it-1}}{f_t} + \lambda_i + \frac{u_{it}}{f_t},$$

we have

$$\frac{y_{it}}{f_t} - \frac{y_{it-1}}{f_{t-1}} = \rho \left(\frac{y_{it-1}}{f_t} - \frac{y_{it-2}}{f_{t-1}} \right) + \frac{u_{it}}{f_t} - \frac{u_{it-1}}{f_{t-1}}.$$

They suggested using lagged variables $\{y_{it-s}\}_{s \geq 2}$ as instruments to estimate ρ and (f_2, \dots, f_T) using the GMM method. This approach may work well when T is small. However, if T is large, this approach becomes problematic because $\{f_2, \dots, f_T\}$ become another set of incidental parameters to be accommodated. One must also address the issue that the number of the instruments then increases at the order $O(T^2)$.⁶

3.3.2 Principal Component Approach with Joint Asymptotics

Another approach is to estimate the model (25) together with $\{\lambda_i' f_t\}_{i,t}$ by least squares and apply joint asymptotics as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2 > 0$ to characterize the form of the bias. The estimate may then be bias corrected using a plug in estimate of the resulting bias.

To fix ideas, let Y, Y_{-1}, U be $N \times T$ matrices whose $(i, t)^{th}$ elements are y_{it}, y_{it-1} , and u_{it} , respectively. Let λ and f be the $(N \times R)$ and the $(T \times R)$ matrices that stack the R - row vectors λ_i' and f_t' , respectively. We may then write the model (25) in matrix notation as

$$Y = \rho Y_{-1} + \lambda f' + U.$$

⁵Pesaran (2006) studied a common correlated random effects model that allows for heterogeneous regression coefficients.

⁶See, for example, Han and Phillips (2006) and Newey and Windmeijer (2009) for a treatment of the problem of many IVs.

Notice that λ and f are not separately identified. One can always transform or rotate with an invertible matrix giving $\tilde{\lambda} = \lambda S$ and $\tilde{f} = f S^{-1}$ such that $\lambda f' = \tilde{\lambda} \tilde{f}'$. However, to identify the parameter of interest ρ , we do not need to separately identify λ and f . Instead, identification of $\lambda f'$ is enough. Details of the identification of ρ are given in Moon and Weidner (2013).

The (negative) Gaussian quasi-loglikelihood function conditioned on the initial conditions $\{y_{i0}\}$ and the interactive fixed effects $\{\lambda'_i f_t\}$ is (up to a constant)

$$l_{NT}(\rho, \lambda, f) = \text{tr} \left[(Y - \rho Y_{-1} - \lambda f')' (Y - \rho Y_{-1} - \lambda f') \right].$$

The profile quasi-loglikelihood function is

$$\begin{aligned} l_{NT}(\rho) &= \min_{\lambda, f} l_{NT}(\rho, \lambda, f) \\ &= \min_f \text{tr} \left[(Y - \rho Y_{-1}) M_f (Y - \rho Y_{-1})' \right] \\ &= \sum_{t=R+1}^T \mu_t \left[((Y - \rho Y_{-1})' (Y - \rho Y_{-1})) \right], \end{aligned}$$

where $M_f = I_T - f(f'f)^{-1}f'$ and $\mu_k[A]$ is the k^{th} smallest eigenvalue of matrix A . The QMLE or the fixed effect estimator $\hat{\rho}$ minimizes $l_{NT}(\rho)$:

$$\hat{\rho} = \arg \min l_{NT}(\rho).$$

Moon and Weidner (2010, 2013) analyzed the properties of this estimator. As before, suppose that $u_{it} \sim$ independent across i and t with a finite uniform 8^{th} moment, that is, $\sup_{i,t} \mathbb{E}(u_{it}^8) < \infty$. Also, assume that λ_i^0 and f_t^0 are R -vector strong factors in the sense that

$$\frac{1}{N} \lambda^{0'} \lambda^0 \rightarrow_p \Sigma_\lambda > 0 \text{ and } \frac{1}{T} f^{0'} f^0 \rightarrow_p \Sigma_f > 0.$$

Under these conditions, they showed that $\hat{\rho}$ is consistent, that is,

$$\hat{\rho} \rightarrow_p \rho_0.$$

Their consistency proof is different from the conventional consistency proof of an extremum estimator that uses a uniform law of large numbers and an identification condition. Moon and Weidner (2013)'s proof is to bound $l_{NT}(\hat{\rho})$ by a lower and an upper bound as follows

$$\begin{aligned} & c(\hat{\rho} - \rho_0)^2 + O_p \left(\frac{|\hat{\rho} - \rho_0|}{\sqrt{\min\{N, T\}}} \right) + \frac{1}{NT} \text{tr}(U'U) + O_p \left(\frac{1}{\min\{N, T\}} \right) \\ & \leq \frac{l_{NT}(\hat{\rho})}{NT} \leq \frac{l_{NT}(\rho_0)}{NT} \\ & = \frac{1}{NT} \sum_{t=R+1}^T \mu_t \left[(\lambda^0 f^{0'} + U)' (\lambda^0 f^{0'} + U) \right] \leq \frac{1}{NT} \text{tr}(U'U) \leq 0, \end{aligned}$$

from which expression and the given rates we can deduce that

$$\hat{\rho} - \rho_0 = O_p \left(\frac{1}{\sqrt{\min\{N, T\}}} \right) = o_p(1).$$

A remaining challenge is to derive the limit distribution of $\hat{\rho} - \rho_0$. The problem is challenging because the conventional approach to deriving the limiting distribution of an extremum estimator typically uses a quadratic approximation of the objective function obtained, for example, via a Taylor approximation. Instead, Moon and Weidner (2010, 2013) use the perturbation theory of a linear operator to approximate $l_{NT}(\rho)$ with a quadratic function of ρ as

$$l_{NT}(\rho) - l_{NT}(\rho_0) = l_{q,NT}(\rho) + \mathcal{R}_{NT}(\beta),$$

where

$$\begin{aligned} l_{q,NT}(\rho) &= -2\sqrt{NT}(\rho - \rho_0) C_{NT} + \left(\sqrt{NT}(\rho - \rho_0) \right)^2 W_{NT} \\ C_{NT} &= C^{(1)}(\lambda^0, f^0, Y_{-1}, U) + C^{(2)}(\lambda^0, f^0, Y_{-1}, U) \\ C^{(1)}(\lambda^0, f^0, Y_{-1}, U) &= \frac{1}{\sqrt{NT}} \text{tr}(M_{f^0} U' M_{\lambda^0} Y_{-1}) \\ C^{(2)}(\lambda^0, f^0, Y_{-1}, U) &= -\frac{1}{\sqrt{NT}} \left[\begin{array}{l} \text{tr} \left(U M_{f^0} U' M_{\lambda^0} Y_{-1} f^0 (f^{0'} f^0)^{-1} (\lambda^{0'} \lambda^0)^{-1} \lambda^{0'} \right) \\ + \text{tr} \left(U' M_{\lambda^0} U M_{f^0} Y_{-1}' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} \right) \\ + \text{tr} \left(U' M_{\lambda^0} Y_{-1} M_{f^0} U' \lambda^0 (\lambda^{0'} \lambda^0)^{-1} (f^{0'} f^0)^{-1} f^{0'} \right) \end{array} \right] \\ W_{NT} &= \frac{1}{NT} \text{tr}(M_{f^0} Y_{-1}' M_{\lambda^0} Y_{-1}), \end{aligned}$$

and

$$\sup_{|\rho - \rho_0| \leq \eta_{NT}} \frac{\mathcal{R}_{NT}(\beta)}{\left(1 + \sqrt{NT}(\rho - \rho_0)^2\right)} = o_p(1),$$

for any sequence $\eta_{NT} \rightarrow 0$. An immediate consequence of the quadratic approximation is that if $C_{NT} = O_p(1)$, then

$$\sqrt{NT}(\hat{\rho} - \rho_0) = W_{NT}^{-1} C_{NT} + o_p(1).$$

Let $v_{it} = \sum_{\tau=0}^{\infty} \rho_0^\tau u_{it-\tau}$ and define the $(N \times T)$ matrix V with v_{it} . Also, let

$F_t^0 = \sum_{\tau=0}^{\infty} \rho_0^\tau f_{t-\tau}^0$ and define the $(T \times R)$ matrix F^0 . Let ⁷

$$\begin{aligned} W &= \lim_{N,T} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}(v_{it-1}^2), \\ \Omega &= \lim_{N,T} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}(u_{it}^2) \mathbb{E}(v_{it-1}^2), \\ B_1 &= p \lim_{N,T} \frac{1}{N} \text{tr} [P_{f^0} \mathbb{E}(U'V)], \\ B_3 &= p \lim_{N,T} \frac{1}{N} \text{tr} [\mathbb{E}(U'U) M_{f^0} F^0 (f^{0'} f^0)^{-1} f^{0'}]. \end{aligned}$$

Suppose that $|\lambda_i^0|$ and $|f_t^0|$ are uniformly bounded across i, t . Then, according to Moon and Weidner (2010), as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2 > 0$,

$$\sqrt{NT}(\hat{\rho} - \rho_0) \Rightarrow N(-\kappa W^{-1}(B_1 + B_3), W^{-1}\Omega W^{-1}).$$

Here the first bias component B_1 arises because the regressor is a lagged dependent variable and so that the regressor is sequentially exogenous, not strictly exogenous (Bai (2009) does not have this bias since that work assumed only strictly exogenous regressors). The second bias component B_3 arises when the error term u_{it} is not homoskedastic. (see Bai (2009) and Moon and Weidner (2010)).

Nonparametric estimators of W, Ω, B_1 , and B_3 are proposed by Moon and Weidner (2010) to achieve bias correction. The resulting bias corrected estimator is

$$\check{\rho}^+ = \hat{\rho} + \hat{W}^{-1} \left(\frac{\hat{B}_1}{T} + \frac{\hat{B}_3}{T} \right),$$

and it is shown that this estimator has the centred limit theory

$$\sqrt{NT}(\check{\rho}^+ - \rho_0) \Rightarrow N(0, W^{-1}\Omega W^{-1}).$$

When the error u_{it} is homoskedastic with $\mathbb{E}(u_{it}^2) = \sigma^2$, we have

$$\sqrt{NT}(\check{\rho}^+ - \rho_0) \Rightarrow N(0, 1 - \rho^2)$$

which is the same distribution as the bias-corrected estimator of Hahn and Kuersteiner (2002) with incidental individual effects.

3.4 Incidental Trends: $\delta(\lambda_i, f_t)$ or $\alpha(\lambda_i, f_t) = \lambda_{i0} + \lambda_{i1}t$

In this section, we turn our attention to panel models with trends. To do so it is convenient to consider a components model specification in which the

⁷The bias term B_2 in Moon and Weidner (2010) is zero when the regressor is a lagged dependent variable.

observed panel data y_{it} consist of a cross-sectionally heterogenous linear trend superposed with serially correlated errors. Our discussion of this model is in two parts, depending on whether the serial correlation in the errors is weak or strong.

3.4.1 Weakly Serially Correlated Case

Suppose that the observed panel y_{it} is generated by the system

$$\begin{aligned} y_{it} &= \lambda_{i0} + \lambda_{i1}t + \varepsilon_{it} \\ \varepsilon_{it} &= \rho\varepsilon_{it-1} + u_{it}, \end{aligned} \tag{26}$$

where $|\rho| < 1$ and $u_{it} \sim iid(0, \sigma^2)$ with $\mathbb{E}|u_{it}|^{4+\varsigma} < \infty$ for some $\varsigma > 0$. In this model the individual time series y_{it} are stationary with deterministic trends. The component panel model (26) can be written in an augmented regression form. Subtracting ρy_{it-1} from y_{it} , we have

$$\begin{aligned} y_{it} &= \lambda_{i0}(1 - \rho) + \lambda_{i1}(t - \rho(t - 1)) + \rho y_{it-1} + u_{it} \\ &= \tilde{\lambda}_{i0} + \tilde{\lambda}_{i1}t + \rho y_{it-1} + u_{it}. \end{aligned} \tag{27}$$

The main difference between (4) and (27) is that the individual effects in (27) are now time varying.

Again, suppose that the object of interest with model (27) is to estimate ρ . Phillips and Sul (2007)⁸ showed that the QMLE estimator of ρ is asymptotically biased for finite T due to the presence of the incidental trends. The bias in the stationary case is

$$\hat{\rho} - \rho \rightarrow_p -2\frac{1+\rho}{T-1} \left(1 + O\left(\frac{1}{T}\right)\right).$$

One way to address the incidental trend problem is to eliminate the incidental trends $\lambda_{i0} + \lambda_{i1}t$ by a double difference transform, instead of first differencing, as

$$\Delta^2 y_{it} = \rho \Delta^2 y_{it-1} + \Delta^2 u_{it}, \tag{28}$$

where $\Delta^2 y_{it} = \Delta(\Delta y_{it}) = y_{it} - 2y_{it-1} + y_{it-2}$, as suggested in Wansbeek and Knapp (1999). We may then use $\{y_{it-s}\}_{s \geq 3}$ as instruments to estimate ρ in (28).

Another approach is to estimate ρ in (27) by least squares and then apply joint asymptotics as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2 > 0$ to derive a limit theory. In this case we can show that

$$\sqrt{NT}(\hat{\rho} - \rho) \Rightarrow N(-2\kappa(1 + \rho), 1 - \rho^2).$$

The bias corrected estimator is then simply

$$\check{\rho}^{++} = \hat{\rho} + \frac{2}{T}(1 + \hat{\rho}) = \frac{T+2}{T}\hat{\rho} + \frac{2}{T}.$$

⁸Phillips and Sul (2007) also proposed a median unbiased estimator.

It follows that as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$, where $0 < \kappa < \infty$,

$$\sqrt{NT} (\hat{\rho}^{++} - \rho) \Rightarrow \mathcal{N}(0, 1 - \rho^2)$$

which is again the same distribution as in Hahn and Kuersteiner (2002).

3.4.2 Strongly Serially Correlated Case

For a panel whose time series has both deterministic trends and stochastic trends, Moon and Phillips (1999, 2000, 2004) and Phillips and Sul (2007) considered the following model:

$$y_{it} = \lambda_{i0} + \lambda_{i1}t + \varepsilon_{it} \quad (29)$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + u_{it}, \quad (30)$$

where

$$\rho = \left(1 - \frac{\theta}{T}\right).$$

In (29), the time series of panel y_{it} consists of cross-sectionally heterogeneous deterministic trends and highly persistent errors (or stochastic trends). The modeling of autoregressive roots as local to unity is common in the analysis of time series (e.g., Phillips, 1987). While it is known that the parameter θ cannot be consistently estimated from a single time series, Moon and Phillips (1999, 2000, 2004) consider estimation possibilities using panel data. For this, they assume that both the cross sectional dimension N and the time dimension T are large such that $\frac{N}{T} \rightarrow 0$. We also assume that $\{u_{it}\}_{i,t}$ are randomly drawn.

First, suppose that ε_{it} is observed (equivalently, that $\lambda_{i0} + \lambda_{i1}t$ is known). Then, we can estimate θ consistently. To see this, consider

$$\hat{\theta} = T(1 - \hat{\rho}),$$

where $\hat{\rho}$ is the least squares estimator of (30):

$$\hat{\rho} = \frac{\sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it-1} \varepsilon_{it}}{\sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it-1}^2}.$$

As $N, T \rightarrow \infty$, since

$$T(\hat{\rho} - \rho) = \frac{\frac{1}{N} \sum_{i=1}^N \frac{1}{T} \sum_{t=1}^T \varepsilon_{it-1} u_{it}}{\frac{1}{N} \sum_{i=1}^N \frac{1}{T^2} \sum_{t=1}^T \varepsilon_{it-1}^2} \rightarrow_p 0,$$

it follows that

$$\hat{\theta} \rightarrow_p \theta$$

contrary to the case where only time series observations are available. An implication is that when panel data are available, using the cross sectional variation, we can estimate strong serial dependence in the data, measured by ρ in the

vicinity of unity, much more accurately than when only a single time series is available.

Now suppose that the true trends are heterogeneous and unknown, so that $\lambda_{i0} + \lambda_{i1}t$ become incidental trends. Let Δ_c be the quasi-difference operator for some local-to-unity parameter c , so that $\Delta_c y_{it} = y_{it} - (1 - \frac{c}{T}) y_{it-1} = \Delta y_{it} + \frac{c}{T} y_{it-1}$. In this section, we shall denote θ as the true localizing coefficient parameter and c as the parameter used in estimation. Then, the Gaussian quasi log-likelihood function conditional on the initial condition y_{i0} is

$$l_{NT}(c, \lambda_1, \dots, \lambda_N) = -\frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \left(\Delta_c y_{it} - \lambda_{i0} \frac{c}{T} - \lambda_{i1} \left(1 + c \frac{t-1}{T} \right) \right)^2.$$

Given c , the MLE for $\hat{\lambda}_i(c)$ is the OLS estimator of $\Delta_c y_{it}$ on $(\frac{c}{T}, 1 + c \frac{t-1}{T})$. Plugging this into $l_{NT}(c, \lambda_1, \dots, \lambda_N)$, we have the concentrated log-likelihood function

$$l_{NT}(c, \hat{\lambda}_1(c), \dots, \hat{\lambda}_N(c)) = -\frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \left(\Delta_c y_{it} - \hat{\lambda}_{i0}(c) \frac{c}{T} - \hat{\lambda}_{i1}(c) \left(1 + c \frac{t-1}{T} \right) \right)^2.$$

The QMLE of θ is

$$\hat{\theta} = \arg \max_c l_{NT}(c, \hat{\lambda}_1(c), \dots, \hat{\lambda}_N(c)).$$

Moon and Phillips (1999) showed that

$$\frac{1}{NT} l_{NT}(c, \hat{\lambda}_1(c), \dots, \hat{\lambda}_N(c)) \rightarrow_p l(c; \theta)$$

for some function $l(c; \theta)$ uniformly in c , and θ does not maximize the limit function $l(c; \theta)$. This implies that the probability limit of $\hat{\theta}$ is not θ . The inconsistency of $\hat{\theta}$ is due to presence of the unknown incidental trends $\lambda_{i0} + \lambda_{i1}t$ in the panel data. Moon and Phillips (1999) called this inconsistency the ‘‘incidental trend’’ problem.

Moon and Phillips (2000, 2004) investigated how to correct for the bias that arises in the presence of the incidental trends in estimating θ in (29). Moon and Phillips (2000) proposed several estimators based on the OLS detrended data. They showed that the estimators are consistent and asymptotically normal when the true parameter $\theta < 0$ but not when $\theta = 0$ (the unit root case). For example, consider the pooled OLS panel estimator $\hat{\theta}^+$ that corrects for the bias due to the time series serial correlation in ε_{it} . Moon and Phillips (2000) showed that

$$\hat{\theta}^+ \rightarrow_p F(\theta),$$

where $F(\theta) \neq \theta$. The inconsistency arises because in the regression

$$\hat{\varepsilon}_{it} = \left(1 - \frac{\hat{\theta}^+}{T} \right) \hat{\varepsilon}_{it-1},$$

where

$$\begin{aligned} \hat{\varepsilon}_{it} &= y_{it} - \hat{\lambda}_{i0} - \hat{\lambda}_{i1}t \\ \begin{pmatrix} \hat{\lambda}_{i0} \\ \hat{\lambda}_{i1} \end{pmatrix} &= \begin{pmatrix} T & \sum_{t=1}^T t \\ \sum_{t=1}^T t & \sum_{t=1}^T t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T y_{it} \\ \sum_{t=1}^T ty_{it} \end{pmatrix}, \end{aligned}$$

(here $\hat{\lambda}_{i0} + \hat{\lambda}_{i1}t$ is the OLS estimator of the incidental trend) the detrended regressor $\hat{\varepsilon}_{it-1}$ ends up being correlated with the error term u_{it} even after correcting for the bias due to the serial correlation in u_{it} .

The first estimator they proposed to resolve this difficulty is to invert the bias function $F(\theta)$ as

$$\tilde{\theta} = F^{-1}(\hat{\theta}^+).$$

Through numerical analysis, Moon and Phillips (2000) showed that $F^{-1}(\bullet)$ is well defined unless $\theta = 0$. A second estimation method they proposed is to correct for the asymptotic bias of $\hat{\theta}^+$ as an approximately linear function of the parameter θ . They showed that both estimators are consistent and asymptotically normal when $\theta < 0$, but these estimators become invalid when $\theta = 0$.

To overcome the problem at $\theta = 0$, Moon and Phillips (2004) considered an estimation method based on two asymptotic moment conditions. The first moment condition was considered in Moon and Phillips (2000). Let $\omega_T(c)$ and $\lambda_T(c)$ be the biases of the score functions of the concentrated likelihoods of the OLS detrended panel and GLS detrended panel, respectively:

$$\begin{aligned} \omega_T(c) &= \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T \left(\hat{\varepsilon}_{it} - \left(1 - \frac{c}{T}\right) \hat{\varepsilon}_{it-1} \right) \hat{\varepsilon}_{it-1} \right] \\ \lambda_T(c) &= \mathbb{E} \left[\frac{1}{T} \sum_{t=1}^T u_{it}(\theta, \hat{\lambda}_i(c)) \varepsilon_{it-1}(\hat{\lambda}_i(c)) \right], \end{aligned}$$

where

$$\begin{aligned} u_{it}(c, \lambda_i) &= \Delta_c y_{it} - \Delta_c \lambda_{i0} - \lambda_{i1} \Delta_c t \\ \varepsilon_{it-1}(\lambda_i) &= y_{it-1} - \lambda_{i0} - \lambda_{i1}(t-1). \end{aligned}$$

The two moment conditions that Moon and Phillips (2004) considered are

$$\begin{aligned} m_{1,iT}(c) &= \frac{1}{T} \sum_{t=1}^T \left(\hat{\varepsilon}_{it} - \left(1 - \frac{c}{T}\right) \hat{\varepsilon}_{it-1} \right) \hat{\varepsilon}_{it-1} - \omega_T(c) \\ m_{2,iT}(c) &= \frac{1}{T} \sum_{t=1}^T u_{it}(c, \hat{\lambda}_i(c)) \varepsilon_{it-1}(\hat{\lambda}_i(c)) - \lambda_T(c), \end{aligned}$$

and the corresponding GMM estimator is

$$\hat{\theta}_{GMM} = \arg \min_{c \leq 0} \left(\frac{1}{N} \sum_{i=1}^N m_{iT}(c) \right)' \bar{W} \left(\frac{1}{N} \sum_{i=1}^N m_{iT}(c) \right),$$

where $m_{iT}(c) = (m_{1,iT}(c), m_{2,iT}(c))'$ and $\hat{W} \rightarrow_p W > 0$. Moon and Phillips (2004) showed that when $\theta < 0$, the GMM estimator $\hat{\theta}_{GMM}$ is \sqrt{N} -consistent and asymptotically normal as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow 0$. When $\theta = 0$, $\hat{\theta}_{GMM}$ is $N^{1/6}$ -consistent and has a nonstandard limiting distribution. So, θ is consistently estimable at $\theta = 0$. However, an important implication of the limit theory is that the presence of incidental trends complicates the identification of a unit root in panels - making it difficult to discriminate locally in the vicinity of unity. This difficulty motivates the investigation of the power of panel unit root tests, as in Moon, Perron, and Phillips (2007).

4 Testing for Unit Roots with Incidental Parameters

A large literature has developed in testing for unit roots in dynamic panels over the past two decades. Information from cross-section observations should be useful in helping to improve inference regarding the long-run properties of data relative to the standard time series tests. However, as the work on local asymptotics described in the previous section makes clear, there are still substantial difficulties in getting good discriminatory power in the immediate vicinity of unity. Our discussion in what follows will focus on the consequences of incidental parameters on the testing problem. Readers should refer to Breitung and Pesaran (2008) for a thorough survey of the area.

We consider a component model where the autoregressive parameter is allowed to be heterogeneous:

$$\begin{aligned} y_{it} &= \lambda_{0i} + \lambda_{1i}t + u_{it} \\ u_{it} &= \rho_i u_{it-1} + \varepsilon_{it}, \end{aligned} \tag{31}$$

where the initial conditions are $y_{i,t-1} = 0$ for all i . For expositional simplicity, we start by assuming that ε_{it} is potentially heteroskedastic with mean 0 and variance σ_i^2 , but that ε_{it} is independent across i and over t with finite fourth moments. The case with serial correlation will be considered later.

The focus of interest is the problem of testing for the presence of a common unit root in the panel when both N and T are large, which we express as the null hypothesis

$$\mathbb{H}_0 : \rho_i = 1 \text{ for all } i, \tag{32}$$

against the alternative

$$\mathbb{H}_1 : \rho_i \neq 1 \text{ for some } i's. \tag{33}$$

It turns out to be convenient to consider a local alternative specification. We assume that

$$\rho(\theta_i) = 1 - \frac{\theta_i}{N^\kappa T} \text{ for some constant } \kappa > 0, \tag{34}$$

where θ_i is a sequence of iid random variables. As we see below, the constant κ will depend on the nature of the incidental parameters.

Moon, Perron, and Phillips (2007) considered efficient tests for the null hypothesis of a unit root for all individuals in the panel. In terms of the local specification, the null and alternative hypotheses can be formulated as:

$$\mathbb{H}_0 : \theta_i = 0 \text{ for all } i, \quad (35)$$

against the alternative

$$\mathbb{H}_1 : \theta_i \neq 0 \text{ for some } i's. \quad (36)$$

If the alternative was a singleton, i.e. if the alternative were some specific value of the local-to-unity parameters which we may call $c = (c_1, c_2, \dots, c_N)'$, then by the Neyman-Pearson lemma, the most powerful test is given by the likelihood ratio. As in Elliott et al. (1996) for the time series case, changing the values of c enables the computation of a power envelope from which we may trace out the maximum power of a test for any point alternative. Under Gaussianity, the log-likelihood is given (up to a constant) by

$$L_{NT}(c) = -\frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \frac{1}{\sigma_i^2} (\Delta_{c_i} y_{it} - \lambda_{i0} (1 - \rho(c_i)) - \lambda_{i1} [t - \rho(c_i)(t-1)])^2, \quad (37)$$

and the test statistic is

$$V_{fe,NT} = 2[L_{NT}(c) - L_{NT}(0)] - \mu(c),$$

where $\mu(c)$ is a centering term that ensures that the statistic has mean 0. When deterministic components are included, we appeal to invariance and use the maximum of the likelihood with respect to these parameters.

One interesting result is that the value of κ that defines local neighborhoods is different according to the specification of the deterministic components. Hence, if $\lambda_{i1} = 0$ and only individual intercepts are present, we find that $\kappa = 1/2$ so that the likelihood ratio test can detect alternatives that converge to the null hypothesis of a panel unit root at the rate $\frac{1}{\sqrt{NT}}$. However, if individual trends are present with $\lambda_{i1} \neq \lambda$ for all i , we find that $\kappa = 1/4$ which means that the maximal power that can be achieved by any test is much lower for a given alternative. Thus, the presence of incidental trends reduces the potential power in discriminating between panels where all individual series have unit roots and panels where some of the individual series have highly persistent but stationary dynamics. Another interesting result is that the presence of incidental intercepts does not change the asymptotic distribution of the test statistic, i.e. the power envelope and distribution of the test statistic is the same whether $\lambda_{i0} = 0$ or $\lambda_{i0} \neq 0$.

As a result, in the incidental intercepts case, Moon et al. (2007) parametrize the autoregressive parameters as:

$$\rho(\theta_i) = 1 - \frac{\theta_i}{\sqrt{NT}},$$

and the null hypothesis can be written as:

$$\mathbb{H}_0 : \mathbb{E}(\theta_i) = 0$$

while the alternative hypothesis is:

$$\mathbb{H}_1 : \mathbb{E}(\theta_i) > 0$$

where the θ_i 's are assumed to be independent across i and lie in the bounded interval $[0, M_\theta]$ for some $M_\theta \geq 0$. The centering term in the likelihood ratio statistic (37) is $\mu(c) = \frac{1}{2N} \sum_{i=1}^N c_i^2$, and its asymptotic distribution is $N(-\mathbb{E}(c_i \theta_i), 2\mu(c))$ which reflects the impact of both the values of the local-to-unity parameters in the population (θ_i) and those used to set up the test and compute the likelihood ratio statistic (c_i).

Moon et al. (2007) make the above test operational by proposing an estimator for σ_i^2 and suggesting a common-point-optimal test in which one chooses all c_i to be the same. In that case, the asymptotic power of a test at level α is $\Phi\left(\frac{\mathbb{E}(\theta_i)}{\sqrt{2}} - \bar{z}_\alpha\right)$ where \bar{z}_α is the $(1 - \alpha)$ quantile from the standard normal distribution. A remarkable feature of this result is that power is independent of the value of the common c_i chosen, in contrast to the time series case where power depends on the choice of the local to unity parameter used to construct the test.

In the incidental trends case, that is when $\lambda_{i1} \neq \lambda$ for all i , as already mentioned, the local neighborhoods must shrink to 1 at a slower rate and the autoregressive parameters are parametrized as (see Ploberger and Phillips, 2002):

$$\rho(\theta_i) = 1 - \frac{\theta_i}{N^{1/4}T}.$$

In this instance, it is possible to allow for some possibly explosive behavior and the assumption made on θ_i is that it is contained in a bounded interval $[-M_{l\theta}, M_{u\theta}]$ where $M_{l\theta}$ and $M_{u\theta}$ are non-negative constants. Under this assumption, the null hypothesis of a panel unit root can be expressed as:

$$\mathbb{H}_0 : \mathbb{E}(\theta_i^2) = 0$$

while the alternative hypothesis is:

$$\mathbb{H}_1 : \mathbb{E}(\theta_i^2) > 0.$$

The use of the second moment is necessary since the requirement that $\mathbb{E}(\theta_i) = 0$ does not imply that all θ_i 's are 0.

Under this scenario, the centering term in the likelihood ratio statistic (37) is

$$\mu(c) = -\frac{1}{N^{1/4}} \sum_{i=1}^N c_i - \frac{1}{N^{1/2}} \sum_{i=1}^N c_i^2 \omega_{p2T} - \frac{1}{N} \sum_{i=1}^N c_i^4 \omega_{p4T},$$

where

$$\begin{aligned}\omega_{p2T} &= -\frac{1}{T} \sum_{t=1}^T \frac{t-1}{T} + \frac{2}{T} \sum_{t=1}^T \frac{t}{T} \left(\frac{t-1}{T} \right) - \frac{1}{3}, \\ \omega_{p4T} &= \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T \frac{t-1}{T} \frac{s-1}{T} \min \left(\frac{t-1}{T}, \frac{s-1}{T} \right) - \frac{2}{3} \frac{1}{T} \sum_{t=1}^T \left(\frac{t-1}{T} \right)^2 + \frac{1}{9}.\end{aligned}$$

The asymptotic distribution is now $N \left(-\frac{1}{90} \mathbb{E} (c_i^2 \theta_i^2), \frac{1}{45} \mathbb{E} (c_i^4) \right)$. A common point optimal test that is constructed by imposing the same value of c_i for all units would have power $\Phi \left(\frac{1}{6\sqrt{5}} \mathbb{E} (\theta_i^2) - \bar{z}_\alpha \right)$, again independent of the particular choice of common local-to-unity parameter used to set up the test. However, this test has power below the power envelope unless the alternative hypothesis is homogeneous, i.e. all θ_i 's are the same. Based on simulation results, Moon et al. (2007) recommend the use of $c_i = 1$ to construct the test.

Moon et al. (2013) extend these results to the case where ε_{it} is serially correlated. They show that optimal tests under serial correlation involve two adjustments to the above statistics. First, the error variance in the denominator must be replaced by the corresponding long-run variance denoted ω_i^2 , and, second, the centering of the statistic must be adjusted to account for the correlation between the error and the lagged dependent variable as in the standard Phillips-Perron statistic in the time series case. This adjustment depends on the one-sided long-run variance $\Lambda_i = \sum_{j=1}^{\infty} \mathbb{E} (\varepsilon_{it} \varepsilon_{t-j})$.

In the case of incidental intercepts, the new centering is $\mu(c) = \frac{1}{2N} \sum_{i=1}^N c_i^2 + \frac{2}{\sqrt{N}} \sum_{i=1}^N c_i \frac{\Lambda_i}{\omega_i^2}$, whereas with incidental trends, the centering is

$$\mu(c) = -\frac{1}{N^{1/4}} \sum_{i=1}^N c_i \frac{\sigma_i^2}{\omega_i^2} - \frac{1}{N^{1/2}} \sum_{i=1}^N c_i^2 \omega_{p2T} - \frac{1}{N} \sum_{i=1}^N c_i^4 \omega_{p4T}.$$

In both cases, the new centering reduces to the centering term in Moon et al. (2007) as $\Lambda_i = 0$ and $\sigma_i^2 = \omega_i^2$ if no serial correlation is present.

5 Nonlinear Dynamic Panels

The incidental parameter problem in nonlinear panel regressions where the incidental parameters are not additively separable is a further challenging problem, in particular, when T is fixed. Surveys of the early literature are available in Arellano and Honore (2001) and Hsiao (2003). For more recent research results, readers may refer to Arellano and Hahn (2006) and Arellano and Bonhomme (2010). This section provides a selective survey on more recent developments of nonlinear dynamic panel regression research. In particular, we focus on estimating the common parameter⁹ of the nonlinear dynamic panel regression model

⁹Another parameter of interest that has been widely studied in nonlinear panel regression models is the average marginal effect (or average treatment effect). See, for example, Fernandez-Val (JoE), Chernazhukov, Fernandez-Val, Hahn, and Newey (2013, Econometrica).

and on bias reduction methods.

5.1 Concentrated Likelihood Approach

Hahn and Newey (2004), Arellano and Hahn (2006), and Hahn and Kuersteiner (2011) extended the idea of Hahn and Kuersteiner (2002) to nonlinear panel regression models and characterize the bias due to the incidental parameters using joint asymptotics as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$. In particular, Hahn and Kuersteiner (2011) allowed the panel regression to be nonlinear and dynamic in which the individual fixed effect parameters and the lagged dependent variables enter the regression model nonlinearly.

Suppose that the observed panel data is x_{it} . The typical composition of x_{it} is $x_{it} = (y_{it}, y_{it-1}, z_{it})$, where y_{it} is the dependent variable in the panel regression and (y_{it-1}, z_{it}) are regressors. Let θ be the common parameters of interest, including the coefficient of the lagged dependent regressor y_{it-1} , and λ_i be the individual fixed effects. Let $(\theta_0, \lambda_1^0, \dots, \lambda_N^0)$ be the true parameters. Consider the fixed effect estimator that maximizes some objective function

$$\left(\hat{\theta}, \hat{\lambda}_1, \dots, \hat{\lambda}_N\right) = \arg \max \sum_{i=1}^N \sum_{t=1}^T \psi(x_{it}; \theta, \lambda_i).$$

Denote

$$l_{it}(\theta, \lambda_i) = \psi(x_{it}; \theta, \lambda_i).$$

An example of $l_{it}(\theta, \lambda_i)$ is the conditional log likelihood function. Define

$$\begin{aligned} w_{it}(\theta, \lambda_i) &= \frac{\partial l_{it}(\theta, \lambda_i)}{\partial \theta}, \quad w_{it} = w_{it}(\theta_0, \lambda_i^0), \quad v_{it}(\theta, \lambda_i) = \frac{\partial l_{it}(\theta, \lambda_i)}{\partial \lambda_i}, \quad v_{it} = v_{it}(\theta_0, \lambda_i^0), \\ w_{it}^\lambda &= \frac{\partial w_{it}(\theta_0, \lambda_i^0)}{\partial \lambda_i}, \quad v_{it}^\lambda = \frac{\partial v_{it}(\theta_0, \lambda_i^0)}{\partial \lambda_i} \\ V_{2it}(\theta, \lambda_i) &= v_{it}^2(\theta, \lambda_i) + \frac{\partial v_{it}(\theta, \lambda_i)}{\partial \lambda_i}, \quad W_{it}(\theta, \lambda_i) = w_{it}(\theta, \lambda_i) - v_{it}(\theta, \lambda_i) \mathbb{E}[v_{it}^\lambda]^{-1} \mathbb{E}[w_{it}^\lambda] \\ W_{it} &= W_{it}(\theta_0, \lambda_i^0), \quad W_{it}^\lambda = \frac{\partial W_{it}(\theta_0, \lambda_i^0)}{\partial \lambda_i}, \quad W_{it}^{\lambda\lambda} = \frac{\partial^2 W_{it}(\theta_0, \lambda_i^0)}{\partial \lambda_i^2}. \\ \mathcal{I}_i &= -\mathbb{E} \left[\frac{\partial W_{it}(\theta_0, \lambda_i^0)}{\partial \theta} \right]. \end{aligned}$$

Hahn and Kuersteiner (2011) showed that in a general nonlinear dynamic panel regression model, as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$,

$$\sqrt{NT} \left(\hat{\theta} - \theta_0 \right) \Rightarrow N \left(\kappa B, \mathcal{I}^{-1} \Omega \mathcal{I}^{-1} \right), \quad (38)$$

where

$$\begin{aligned}\mathcal{I} &= \lim_N \frac{1}{N} \sum_{i=1}^N \mathcal{I}_i, \quad \Omega = \lim_N \frac{1}{N} \sum_{i=1}^N \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T W_{it} \right) \\ B &= -\mathcal{I}^{-1} \left(\lim_N \frac{1}{N} \sum_{i=1}^N \frac{\sum_{l=-\infty}^{\infty} \text{Cov}(v_{it}, W_{it-l}^\lambda)}{\mathbb{E}[v_{it}^\lambda]} - \frac{1}{2} \lim_N \frac{1}{N} \sum_{i=1}^N \frac{\mathbb{E}[W_{it}^{\lambda\lambda}] \sum_{l=-\infty}^{\infty} \text{Cov}(v_{it}, v_{it-l})}{\mathbb{E}[v_{it}^\lambda]^2} \right).\end{aligned}$$

Notice that in the dynamic linear regression model (4), we have

$$B = \frac{-(1-\rho_0)}{1-\rho_0^2} = -\frac{1}{1+\rho_0},$$

as shown in (20). Hahn and Kuersteiner (2011) also provide a consistent estimator \hat{B} of the bias B , and propose a bias corrected estimator

$$\hat{\theta}^+ = \hat{\theta} - \frac{1}{T} \hat{B}.$$

The intuition underlying (38) is as follows. Consider an infeasible estimator $\tilde{\theta}$ based on $\hat{\lambda}_i(\theta_0)$ rather than $\hat{\lambda}_i(\tilde{\theta})$, where

$$0 = \sum_{i=1}^N \sum_{t=1}^T W_{it}(\tilde{\theta}, \hat{\lambda}_i(\theta_0)).$$

Then, conventional first order Taylor approximation yields

$$\tilde{\theta} - \theta_0 \simeq \left(\frac{1}{N} \sum_{i=1}^N \mathcal{I}_i \right)^{-1} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T W_{it}(\theta_0, \hat{\lambda}_i(\theta_0)).$$

Applying a second order Taylor series approximation to $W_{it}(\theta_0, \hat{\lambda}_i(\theta_0))$, we have

$$\begin{aligned}\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T W_{it}(\theta_0, \hat{\lambda}_i(\theta_0)) &\simeq \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T W_{it} + \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T W_{it}^\lambda (\hat{\lambda}_i(\theta_0) - \lambda_i^0) \\ &\quad + \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T W_{it}^{\lambda\lambda} (\hat{\lambda}_i(\theta_0) - \lambda_i^0)^2.\end{aligned}$$

Since $\hat{\lambda}_i(\theta_0) - \lambda_i^0 \simeq -\frac{1}{T} \sum_{t=1}^T v_{it} (\mathbb{E}(v_{it}^\lambda))^{-1}$, we have

$$\begin{aligned}\sqrt{NT}(\tilde{\theta} - \theta_0) &\simeq \left(\frac{1}{N} \sum_{i=1}^N \mathcal{I}_i \right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T W_{it} \right) \\ &\quad - \sqrt{\frac{N}{T}} \left(\frac{1}{N} \sum_{i=1}^N \mathcal{I}_i \right)^{-1} \frac{1}{N} \sum_{i=1}^N \left[\frac{\sum_{t=1}^T v_{it}}{\sqrt{T} \mathbb{E}(v_{it}^\lambda)} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(W_{it}^\lambda - \frac{\mathbb{E}[W_{it}^{\lambda\lambda}]}{2\mathbb{E}(v_{it}^\lambda)} v_{it} \right) \right],\end{aligned}$$

and it follows that

$$\left(\frac{1}{N} \sum_{i=1}^N \mathcal{I}_i \right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T W_{it} \right) \Rightarrow N(0, \mathcal{I}^{-1} \Omega \mathcal{I}^{-1})$$

and

$$-\sqrt{\frac{N}{T}} \left(\frac{1}{N} \sum_{i=1}^N \mathcal{I}_i \right)^{-1} \frac{1}{N} \sum_{i=1}^N \left[\frac{\sum_{t=1}^T v_{it}}{\sqrt{T} \mathbb{E}(v_{it}^\lambda)} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(W_{it}^\lambda - \frac{\mathbb{E}[W_{it}^{\lambda\lambda}]}{2\mathbb{E}(v_{it}^\lambda)} v_{it} \right) \right] \rightarrow_p \kappa B.$$

5.2 Integrated Likelihood Approach

The fixed effect estimator concentrates out the incidental parameters in the objective function. Another way to deal with the incidental parameters is to integrate out with a certain weight function or a prior for the parameters. This approach was studied by Arellano and Bonhomme (2009).

Using the notation in this section, suppose that $l_{it}(\theta, \lambda)$ is the conditional log likelihood function. Let $l_i(\theta, \lambda) = \frac{1}{T} \sum_{t=1}^T \ln l_{it}(\theta, \lambda)$. Let $\pi_i(\lambda_i|\theta)$ be a conditional prior distribution on the individual fixed effect given θ . Here the dependence of the π_i on the index i allows for possible conditioning on strictly exogenous regressors and initial conditions. Assume that the support of $\pi_i(\lambda_i|\theta)$ contains an open neighborhood of the true parameter (θ_0, λ_i^0) and $\sup_i \ln \pi_i(\lambda_i|\theta) = O(1)$ for all θ and λ_i as $T \rightarrow \infty$.

The fixed effect estimator $\hat{\theta}$ maximizes the concentrated likelihood function

$$\hat{\theta} = \arg \max \sum_{i=1}^N l_i^c(\theta),$$

where

$$l_i^c(\theta) = l_i(\theta, \hat{\lambda}_i(\theta)).$$

An alternative objective function to the concentrated likelihood is the individual log integrated likelihood given by

$$l_i^I(\theta) = \frac{1}{T} \ln \int \exp[Tl_i(\theta, \lambda_i)] \pi_i(\lambda_i|\theta) d\lambda_i.$$

This likelihood could be considered subjective Bayesian with a joint prior that is separable in the individual effects. The target likelihood is defined as

$$\bar{l}_i(\theta) = l_i(\theta, \bar{\lambda}_i(\theta)),$$

where

$$\bar{\lambda}_i(\theta) = \arg \max_{\lambda_i} p \lim_T l_i(\theta, \lambda_i).$$

Notice that the concentrated and target likelihood functions can be regarded as integrated likelihood functions with respect to the priors

$$\bar{\pi}_i(\lambda_i|\theta) = \delta(\lambda_i - \bar{\lambda}_i(\theta)) \quad \text{and} \quad \pi_i^c(\lambda_i|\theta) = \delta(\lambda_i - \hat{\lambda}_i(\theta)),$$

where $\delta(\cdot)$ is the Dirac delta function. Here $\pi_i^c(\lambda_i|\theta)$ is a sample counterpart of $\bar{\pi}_i(\lambda_i|\theta)$. Define $v_i(\theta, \lambda_i) = \frac{\partial l_i(\theta, \lambda_i)}{\partial \lambda_i}$ and $v_i^{\lambda_i}(\theta, \lambda_i) = \frac{\partial v_i(\theta, \lambda_i)}{\partial \lambda_i}$, $v_i^\theta(\theta, \lambda_i) = \frac{\partial v_i(\theta, \lambda_i)}{\partial \theta}$, and $v_i^{\lambda_i \lambda_i}(\theta, \lambda_i) = \frac{\partial^2 v_i(\theta, \lambda_i)}{\partial \lambda_i^2}$.

Arellano and Bonhomme (2009) showed that the bias of the integrated likelihood is

$$\mathbb{E} [l_i^I(\theta) - \bar{l}_i(\theta)] = \text{const} + \frac{\beta_i(\theta)}{T} + O\left(\frac{1}{T^2}\right),$$

where

$$\beta_i(\theta) = \frac{1}{2} \left\{ \mathbb{E} \left[-v_i^{\lambda_i}(\theta, \bar{\lambda}_i(\theta)) \right] \right\}^{-1} \mathbb{E} [T v_i^2(\theta, \bar{\lambda}_i(\theta))] - \frac{1}{2} \ln \mathbb{E} \left[-v_i^{\lambda_i}(\theta, \bar{\lambda}_i(\theta)) \right] + \ln \pi_i(\bar{\alpha}_i(\theta) | \theta).$$

Let $b_i(\theta_0) = \frac{\partial}{\partial \theta} |_{\theta_0} \beta_i(\theta)$ be the first order bias of the integrated score evaluated at the true value. Arellano and Bonhomme defined a prior family as bias reducing or robust, if and only if

$$b_\infty(\theta_0) = p \lim_N \frac{1}{N} \sum_{i=1}^N b_i(\theta_0) = o(1).$$

Since bias reduction of the moment equation (score function) implies bias reduction of the estimator, for a robust prior family, the mode of the integrated likelihood

$$\hat{\theta}_I = \arg \max_{\theta} \sum_{i=1}^N l_i^I(\theta)$$

has zero first-order bias, that is,

$$p \lim_{N \rightarrow \infty} \hat{\theta}_I = \theta_0 + o\left(\frac{1}{T}\right).$$

Arellano and Bonhomme showed that a prior π_i is bias-reducing if

$$\frac{\partial}{\partial \theta} \Big|_{\theta_0} \ln \pi_i(\bar{\alpha}_i(\theta) | \theta) = \frac{\partial}{\partial \theta} \Big|_{\theta_0} \ln \left(\mathbb{E} \left[-v_i^{\lambda_i}(\theta, \bar{\lambda}_i(\theta)) \right] \left\{ \mathbb{E} [T v_i^2(\theta, \bar{\lambda}_i(\theta))] \right\}^{-1/2} \right) + O\left(\frac{1}{T}\right).$$

They suggested the following data-dependent prior:

$$\pi_i^R(\lambda_i | \theta) \propto \mathbb{E} \left[-\widehat{v_i^{\lambda_i}}(\theta, \lambda_i) \right] \left\{ \mathbb{E} [\widehat{v_i^2}(\theta, \lambda_i)] \right\}^{-1/2},$$

where the hat denotes consistent estimators as $T \rightarrow \infty$. In the pseudo likelihood setting, they suggested

$$\pi_i^R(\lambda_i | \theta) \propto \mathbb{E} \left[-\widehat{v_i^{\lambda_i}}(\theta, \lambda_i) \right]^{1/2} \exp \left(-\frac{T}{2} \left\{ \mathbb{E} \left[-\widehat{v_i^{\lambda_i}}(\theta, \lambda_i) \right] \right\}^{-1} \mathbb{E} [\widehat{v_i^2}(\theta, \lambda_i)] \right).$$

Under these robust priors, as $N, T \rightarrow \infty$ with $\frac{N}{T} \rightarrow \kappa^2$,

$$\sqrt{NT} \left(\hat{\theta}_I - \bar{\theta} \right) = o_p(1),$$

where $\bar{\theta} = \arg \max_{\theta} \sum_{i=1}^N \bar{l}_i(\theta)$. Therefore,

$$\sqrt{NT}(\hat{\theta}_I - \theta_0) = \sqrt{NT}(\bar{\theta} - \theta_0) + o_p(1).$$

6 Order Selection in Dynamic Panels

As indicated in the Introduction, the presence of incidental parameters further complicates model selection. Standard procedures such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or Hannan-Quinn (HQ) suppose that the number of parameters is finite or grows slowly as sample size increases. The presence of incidental parameters violates this assumption, and Stone (1979) showed that BIC is inconsistent with incidental parameters. The same problem occurs in dynamic panels with fixed effects. The reason why BIC is inconsistent in this context is that the usual Laplace approximation does not hold for an infinity of parameters, thereby requiring an infinite dimensional integration.

It is particularly remarkable that BIC is also inconsistent as an order estimator in dynamic panels without fixed effects as shown recently by Han, Phillips and Sul (2012b). The reason for this curious finding is that in a panel as $N \rightarrow \infty$ there are an infinity of observations relevant to models with a lower lag order than those with higher lag order. It is this discrepancy in the use of the data that can lead to inconsistency in conventional order selectors like BIC. Han Phillips and Sul show that the overestimation probability in the BIC order selector is 50%. The remedy is to raise the penalty in the BIC criterion to take account of this difference.

How to perform model selection with incidental parameters remains an ongoing issue in the panel literature although some approaches have been developed recently. Berger et al. (2003) have shown that the choice of priors is important with incidental parameters, and that a suitable choice can make BIC consistent in this context. They also propose a different approximation that leads to a criterion that is consistent in model selection.

Alternatively, in a Kullback-Leibler approach, one must increase the penalty used because standard methods impose penalties that are too small if the problem of incidental parameters is not taken into account. For example, Lee (2012) considers the problem of selecting a model among a set of candidate models that may not contain the true one. Suppose that there are two sets of parameters, those of interest denoted by ψ_k with dimension r_k for model k and the incidental parameters λ_i . He defines a set of information criteria for model k of the form:

$$LIC^h(\mathcal{M}^k) = -\frac{2}{NT} \sum_{i=1}^N \sum_{t=1}^T \log f_{it}(z_{it}; \hat{\psi}_M^k, \hat{\lambda}_i^k) + r_k \frac{h(N, T)}{NT} + \frac{2}{NT} \sum_{i=1}^N M_i(\hat{\psi}_M^k)$$

where $h(N, T)$ is a penalty function that differentiates the criteria. The estimator $\hat{\psi}_M^k$ is based on the modified profile likelihood that corrects for the fact that the score of the profile likelihood is not 0 due to the incidental parameters.

The penalty in these criteria has two components. The first one is the same as in the standard AIC and BIC and is proportional to the number of parameters. For the AIC, one would set $h(N, T) = 2$ while for BIC, one would set $h(N, T) = \log(NT)$. The second term is the contribution associated with the presence of incidental parameters. This term is always positive so that a more severe penalty is imposed relative to the standard criteria.

7 Conclusions

Practical empirical work with dynamic panel models offers many opportunities for learning about individual behavior over time and the common elements that figure in that behavior. This work also faces many challenges, ranging from the impact of individual effects and incidental trends on estimation bias in short wide panels, through to the difficulties of treating nonlinear dynamic models with nonseparable fixed effects, and the problems of inconsistency in dynamic model specification. This chapter has overviewed some of the established methodology in the field, the ground that has been won in developing a theory of inference for dynamic panel modeling, as well as some exciting ongoing research that seeks to address some of the many challenges that remain.

References

- [1] Ahn, S. C., Lee, Y.H., and P. Schmidt (2001): GMM Estimation of Linear Panel Data Models with Time-varying Individual Effects, *Journal of Econometrics*, 101, 219-255.
- [2] Ahn, S., and P. Schmidt (1995): Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-28.
- [3] Anderson, T.W. and C. Hsiao (1982): Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics* 18, 47-82,
- [4] 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.
- [5] Arellano, M., and S. Bonhomme (2009): Robust Priors in Nonlinear Panel Data Models, *Econometrica*, 77, 489-539.
- [6] Arellano, M., and S. Bonhomme (2010): Nonlinear Panel Data Analysis, Working Paper, CEMFI.
- [7] Arellano, M., and O. Bover (1995): Another Look at the Instrumental Variable Estimation of Error-Component Models, *Journal of Econometrics*, 68, 29-51.

- [8] Arellano, M. and J. Hahn (2006): Understanding Bias in Nonlinear Panel Models: Some Recent Developments, in Blundell, R. W.K.Newey, and T. Persson eds. *Advances in Economics and Econometrics*, Cambridge: Cambridge University Press, 381-409.
- [9] Arellano, M., and B. Honore (2001): Panel Data Models: Some Recent Development in Heckman, J.J. and Leamer, E.E. eds., *Handbook of Econometrics*, vol. 5, Elsevier, 3229-3296.
- [10] Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica*, 67, 1341–1384.
- [11] Baltagi, B.H (2008): *Econometric Analysis of Panel Data (Fourth Edition)*, John Wiley and Sons.
- [12] Berger, J. O., Ghosh, J. K. and N. Mukhopadhyay, (2003). Approximations and Consistency of Bayes Factors as the Model Dimension Grows. *Journal of Statistical Planning and Inference*, 112, 241-258.
- [13] Blundell, R. and S. Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.
- [14] Breitung, J. and M. H, Pesaran (2008): Unit Roots and Cointegration in Panels, forthcoming in Matyas, L. and P. Sevestre, *The Econometrics of Panel Data (Third Edition)*, Kluwer Academic Publishers.
- [15] Casella, G., F. J. Girón, M. L. Martinez, M. L. and E. Moreno. (2009). Consistency of Bayesian Procedures for Variable Selection. *Annals of Statistics*, 37, 1207–1228.
- [16] Casella and Moreno, (2006): Objective Bayesian variable selection. *Journal of the American Statistical Association*, 101 157–167.
- [17] Casella G. and E. Moreno (2009). Assessing robustness of intrinsic test of independence in two-way contingency tables. *Journal of the American Statistical Association*, 104 1261–1271.
- [18] Chamberlain, G. (1982): Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18, 5-46.
- [19] Chamberlain, G. (1984): Panel Data, in Griliches Z. and M. Intriligator eds., *Handbook of Econometrics, Volume 2*, Amsterdam: North-Holland, 1247–1318.
- [20] Dhaene, G. and K. Jochmans (2012): Split-Panel Jackknife Estimation of Fixed-Effect Models, Working Paper.
- [21] Elliott, G. T. Rothenberg, and J. Stock (1996): Efficient Tests for an Autoregressive Unit Root, *Econometrica*, 64, 813–836.

- [22] Gouriéroux, C., P.C.B. Phillips, and J. Yu (2010): Indirect Inference for Dynamic Panel Models, *Journal of Econometrics*, 157, 68-77.
- [23] Hahn, J. and G. M. Kuersteiner (2002): Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T are Large, *Econometrica*, 70, 1639-1657.
- [24] Hahn, J. and G. M. Kuersteiner (2011): Bias Reduction for Dynamic Non-linear Panel Models with Fixed Effects, *Econometric Theory*, 27,1152-1191.
- [25] Hahn, J. and H. R. Moon (2006): Reducing Bias of MLE in a Dynamic Panel Model, *Econometric Theory*, 22, 499-512.
- [26] Hahn, J. and W. Newey (2004): Jackknife and Analytical Bias Reduction for Nonlinear Panel Models, *Econometrica*, 72, 1295-1319.
- [27] Han, C. and P.C.B. Phillips (2006): GMM with Many Moment Conditions, *Econometrica*, 74, 147-192.
- [28] Han, C. and P.C.B. Phillips (2010): GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity, *Econometric Theory*, 26, 119-151.
- [29] Han, C., Phillips, P. C. B. and D. Sul. (2011). Uniform Asymptotic Normality in Stationary and Unit Root Autoregression. *Econometric Theory* 27:1117–1151.
- [30] Han, C., Phillips, P. C. B. and D. Sul (2012a): X-Differencing and Dynamic Panel Model Estimation. forthcoming in *Econometric Theory*.
- [31] Han, C., P.C.B. Phillips and D. Sul (2012b): Lag Length Selection in Panel Autoregression. Working paper, Yale University.
- [32] Holtz-Eakin, D., W. Newey, and H. Rosen (1988), Estimating Vector Autoregressions with Panel Data, *Econometrica*, 56, 1371-1395.
- [33] Hsiao, C. (2003). Analysis of Panel Data, Cambridge University Press.
- [34] Kiefer, N. (1980). A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure, Mimeo, Department of Economics, Cornell University.
- [35] Lancaster, T. (2000): The Incidental Parameter Problem Since 1948, *Journal of Econometrics*, 95, 391-413.
- [36] Lee, Y.H. (1991). Panel Data Models with Multiplicative Individual and Time Effects: Application to Compensation and Frontier Production Functions, Ph.D. Dissertation, Michigan State University.
- [37] Lee, Y. (2012). Model selection in the presence of incidental parameters. Unpublished manuscript, University of Michigan.

- [38] Moon, H. R., B. Perron, and P.C.B. Phillips (2007): Incidental Trends and the Power of Panel Unit Root Tests, *Journal of Econometrics*, 141, 416–459.
- [39] Moon, H. R., B. Perron, and P.C.B. Phillips (2013): Point Optimal Panel Unit Root Tests with Serially Correlated Errors, Mimeo.
- [40] Moon, H. R. and P.C.B. Phillips (1999): Maximum Likelihood Estimation in Panels with Incidental Trends, *Oxford Bulletin of Economics and Statistics*, 61, 711-747.
- [41] Moon, H. R. and P.C.B. Phillips (2000): Estimation of Autogressive Roots near Unity using Panel Data, *Econometric Theory*, 16, 927–997.
- [42] Moon, H. R. and P.C.B. Phillips (2004): GMM Estimation of Autoregressive Roots Near Unity with Panel Data, *Econometrica*, 72, 467–522.
- [43] Moon, H.R. and M. Weidner (2010): Dynamic Linear Panel Regression Models with Interactive Fixed Effects, Working Paper.
- [44] Moon, H.R. and M. Weidner (2013): Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects, Working Paper.
- [45] Moreno, E. F. J. Girón and G. Casella (2010): Consistency of Objective Bayes Factors as the Model Dimension Grows, *Annals of Statistics*, 38, 4, 1937-1952.
- [46] Neyman J., and E. Scott (1948): Consistent Estimates Based on Partially Consistent Observations, *Econometrica*, 16, 1–31.
- [47] Nickell, S. (1981): Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1417-1426.
- [48] Pesaran, M.H. (2006). Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica*, 74, 967–1012.
- [49] Phillips, P.C.B. (1987): Towards a Unified Asymptotic Theory for Autoregression, *Biometrika*, 74, 535-547.
- [50] Phillips, P.C.B. and H.R. Moon (1999): Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67, 1057–1111.
- [51] Phillips, P.C.B. and D. Sul (2003): Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence, *Econometrics Journal*, 6, 217–259.
- [52] Phillips, P.C.B. and D. Sul (2007): Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence, *Journal of Econometrics*, 137, 162–188.
- [53] Ploberger, W. and P.C.B. Phillips (2002): Optimal Testing for Unit Roots in Panel Data. Unpublished working paper.

- [54] Stone, M. (1979): Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B*, 41, 276-278.
- [55] Wansbeek, T. and T. Knapp (1999): Estimating a Dynamic Panel Data Model with Heterogeneous Trends, *Annales d'Économie et de Statistique*, 55/56, 331-349.