Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 230–244
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Panel data unit roots tests: The role of serial correlation and the time dimension

Stefan De Wachter[a,*], Richard D.F. Harris[b], Elias Tzavalis[c]

[a] *Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK*
[b] *Xfi Centre for Finance and Investment, University of Exeter, Exeter EX4 4ST, UK*
[c] *Department of Economics, Athens University of Economics & Business, Patission 76, Athens 104 34, Greece*

Received 15 November 2005; accepted 16 November 2005
Available online 4 January 2006

Abstract

We investigate the influence of residual serial correlation and of the time dimension on statistical inference for a unit root in dynamic longitudinal data, known as panel data in econometrics. To this end, we introduce two test statistics based on method of moments estimators. The first is based on the generalized method of moments estimators, while the second is based on the instrumental variables estimator. Analytical results for the Instrumental Variables (IV) based test in a simplified setting show that (i) large time dimension panel unit root tests will suffer from serious size distortions in finite samples, even for samples that would normally be considered large in practice, and (ii) negative serial correlation in the error terms of the panel reduces the power of the unit root tests, possibly up to a point where the test becomes biased. However, near the unit root the test is shown to have power against a wide range of alternatives. These findings are confirmed in a more general set-up through a series of Monte Carlo experiments.

MSC: C22; C23

Keywords: Dynamic longitudinal (panel) data; Generalized method of moments; Instrumental variables; Unit roots; Moving average errors

1. Introduction

There has been much recent interest—both theoretical and applied—in testing for unit roots in longitudinal data, known in econometrics as panel data. Existing panel data unit root tests can be classified into two categories: the first treats the time dimension of the panel, T, as large (see Levin et al., 2002; Im et al., 1995; Hadri, 2000, inter alia) while the second treats T as fixed (short) (see Harris and Tzavalis, 1999). Asymptotic theories for both categories of tests assume that the cross-section dimension of the panel, N, goes to infinity; for large-T tests, also T is assumed to increase without bound, either jointly with N or sequentially. The fixed-T tests can be thought of as more appropriate for panels where the time dimension is small, while the large T tests are naturally suited to those panels where the time dimension can be considered large (see Chamberlain, 1984, inter alia). From the point of view of statistical inference, however, there is no rule that allows one to classify the time dimension of a panel as small or large.

* Corresponding author.
E-mail address: stefan.dewachter@economics.ox.ac.uk (S. De Wachter).

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2005.11.004

In this paper, we assess the influence of the time dimension of a panel on statistical inference for a unit root in the presence of serially correlated errors, a set-up which has proven to be challenging for single time series tests (see Schwert, 1989; Wu and Yin, 1999). Our aims are twofold. Firstly, the paper intends to characterize the specific problems for unit root testing that are introduced by allowing for serial correlation in the errors. We will show that tests lose power as serial correlation grows large and negative, up to a point where they may become biased. Near the unit root, however, the tests are powerful against a wide range of alternatives. Secondly, our study will help to investigate the minimum number of time series observations that are appropriate in order for short or large panel data unit root tests to be applicable. This will shed light on existing evidence that, for both single time series and panel data, unit root test statistics that assume that T is large seem to be critically oversized in small samples, especially when the panel disturbance (error) terms are negatively serially correlated.

We start our study by introducing a Generalized Method of Moments (GMM) based unit root test statistic. This is primarily designed for short T panels, as it is based on cross-sectional averaging only and allows the nuisance parameters to be heterogeneous across both the N and T dimensions of the panel. The paper then introduces an Instrumental Variables (IV) based test statistic under the additional assumption that the nuisance parameters of the panel are homogeneous across both dimensions of the panel. This test can be applied to panels where the T dimension is short, or large. The IV based test statistic will help us to analytically examine the influence of the time dimension of the panel and the serial correlation nuisance parameters on panel data unit root tests for the class of the method of moments based test statistics to which the IV based test belongs. Moreover, it allows us to examine power properties in a tractable setting.

To address these issues in a simple framework, consider the first order autoregressive panel data model,

$$z_{i,t} = \eta_i(1-\rho) + \rho z_{i,t-1} + u_{i,t} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{1}$$

where the error terms $u_{i,t}$ are zero mean p-dependent processes[1] which are independent across i, with $E|u_{i,t}|^{4+\delta}$ uniformly bounded over i and t, for some $\delta > 0$ and $p < T$. η_i are individual-specific long-run means[2] of the processes when $\rho < 1$ in the limit as $\rho \to 1$ the processes become driftless random walks; in this way individual-specific trends are ruled out for all values of ρ. The assumption that $p < T$ means that the order of serial correlation is smaller than the T dimension of the panel. It is required to derive unit root test statistics where T is fixed. The above assumption on the disturbance terms $u_{i,t}$ is quite general yet enables us to apply standard asymptotic results across the N dimension of the panel. When discussing the IV statistic, we will focus on the case where $u_{i,t}$ follows a homogeneous MA(1) process. More specifically

$$u_{i,t} = v_{i,t} + \theta v_{i,t-1}, \tag{2}$$

where the error terms $v_{i,t}$ are independent zero mean random variables with $E|v_{i,t}|^{4+\delta} < \infty$.

At this point we do not make any assumption on the initial conditions of the panel $z_{i,0}$. The test statistics we derive are invariant to $z_{i,0}$ under the null hypothesis $\rho = 1$. This is achieved by subtracting $z_{i,0}$ from each observation $z_{i,t}$ as in Breitung and Meyer (1994): we define the new series

$$y_{i,t} = z_{i,t} - z_{i,0} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{3}$$

and employ $y_{i,t}$ rather than $z_{i,t}$ in deriving the limiting distributions of the test-statistic of the hypothesis $\rho = 1$ in model (1). Model (1) is written in terms of y as

$$y_{i,t} = (\eta_i - z_{i,0})(1-\rho) + \rho y_{i,t-1} + u_{i,t} \quad i = 1, \ldots, N; \quad t = 2, \ldots, T. \tag{4}$$

The paper is organized as follows: Section 2 introduces the test statistics and derives their limiting distribution. Section 3 conducts a Monte Carlo study to appraise the small sample performance of our test statistics and to confirm some of the theoretical results derived in Section 2. Section 4 concludes the paper. All proofs are relegated to the Appendix.

[1] That is $u_{i,t}$ and $u_{i,t-p-1}$ are independent random variables, but $u_{i,t}$ and $u_{i,\tau}$, $t \geqslant \tau \geqslant t-p$ may be dependent. Note that the dependence structure is allowed to be heterogeneous across i.

[2] $\eta_i(1-\rho)$ are often referred to as "fixed effects" in the econometrics literature.