UNIT ROOTS: IDENTIFICATION AND TESTING IN MICRO PANELS

Stephen Bond
Céline Nauges
Frank Windmeijer

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP07/05

Unit Roots: Identification and Testing in Micro Panels*

Stephen Bond
Institute for Fiscal Studies and Nuffield College, Oxford

Céline Nauges
LERNA-INRA, Toulouse

Frank Windmeijer
Centre for Microdata Methods and Practice
Institute for Fiscal Studies, London

July 2005

Abstract

We consider a number of unit root tests for micro panels where the number of individuals is typically large, but the number of time periods is often very small. As we discuss, the presence of a unit root is closely related to the identification of parameters of interest in this context. Calculations of asymptotic local power and Monte Carlo evidence indicate that two simple t-tests based on ordinary least squares estimators perform particularly well.

JEL Classification: C12, C23

Key Words: Generalised Method of Moments, Identification, Unit Root Tests

* **Acknowledgments:** We would like to thank Min Ahn, Manuel Arellano, Bronwyn Hall and Stefan de Wachter for helpful comments. Financial support from the ESRC Centre for Public Policy at IFS and the Leverhulme Trust are gratefully acknowledged. The usual disclaimer applies.

1. Introduction

Microeconomic panel data sets - for example, on individuals or households, or on plants or firms - commonly have a cross-section dimension (N) that is large and a time dimension (T) that is small. Because asymptotic approximations treat the number of time periods as fixed, the presence of non-stationary integrated series does not change the nature of asymptotic distribution results in the same way that it does for single time series or for panels with large T. However, testing for unit roots in micro panels is motivated by considering the properties of several well-known estimators of autoregressive models in the unit root case. Some of these do not identify the parameter of interest in the unit root case, so that evidence on the time series properties of the data may be crucial for the choice of estimator to be considered. There are also economic contexts in which testing the time series properties of microeconomic series is of primary importance, as for example in the empirical literature on firm size and Gibrat's Law.[1]

While there are estimators that are consistent both under the null hypothesis of a unit root and under stationary alternatives, we stress that consistent tests of the unit root hypothesis require consistent estimation only under the null. We show that simple Wald tests based on ordinary least squares estimators can have significantly better power properties than alternative tests that have been suggested in this context.

Section 2 outlines the model we consider, and discusses (under)identification in the unit root case for both Generalised Method of Moments (GMM) and Maximum Likelihood (ML) estimators. Section 3 reviews panel unit root tests that have been

[1] See, for example, Sutton (1997).

considered for panels with large N and fixed T. Section 4 presents asymptotic local power comparisons and section 5 presents Monte Carlo evidence on small sample properties. Section 6 concludes.

2. Model, Estimators and Identification

Consider the simple dynamic AR(1) panel data model

$$\begin{aligned} y_{i1} &= \delta_0 + \delta_1 \eta_i + \varepsilon_i \\ y_{it} &= \alpha y_{i,t-1} + u_{it} \\ u_{it} &= (1-\alpha)\,\eta_i + v_{it}, \end{aligned} \tag{2.1}$$

for $i = 1, ..., N$ and $t = 2, ..., T$, where N is large and T is fixed. The observations are independent across individuals and the error term satisfies

$$E(\eta_i) = 0, \; E(v_{it}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2..., T$$

and

$$E(v_{it} v_{is}) = 0 \quad \text{for } i = 1, ..., N \text{ and } t \neq s.$$

We focus here on tests of the null hypothesis that the series have a unit root ($\alpha = 1$) or are integrated of order one against the alternative that the series are 'stationary' in the sense of being integrated of order zero ($\alpha < 1$). Because the number of time periods considered is small, properties of the initial conditions (y_{i1}) are also relevant for 'stationarity' properties of the series. Mean stationarity (constant first moment) requires $\alpha < 1$ and $\delta_0 = 0$ and $\delta_1 = 1$. Covariance stationarity (constant first and second moments) further requires homoskedasticity over time of the v_{it} shocks (i.e. $Var(v_{it}) = \sigma^2_{vi}$ for $i = 1, ..., N$) and that $Var(\varepsilon_i) = \sigma^2_{vi} / (1 - \alpha^2)$.

This setting is similar to that studied by Breitung and Meyer (1994), Harris and Tzavalis (1999) and Hall and Mairesse (2005). Notice that there are no individual effects in this specification when $\alpha = 1$, so the null hypothesis is that the y_{it} series are random walks with no drifts for each individual. Individual-specific trends are thus ruled out under both the null and the alternative.

2.1. GMM

2.1.1. First-Differenced GMM

If it is only assumed that the y_{i1} are uncorrelated with v_{it}:

$$E\left(y_{i1}v_{it}\right) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2..., T,$$

then there are the following $(T-1)(T-2)/2$ linear moment conditions available for the estimation of α by GMM

$$E\left(y_{is}\Delta u_{it}\right) = 0 \quad \text{for } t = 3, \dots T,\ s = 1, \dots, t-2, \tag{2.2}$$

where $\Delta u_{it} = u_{it} - u_{i,t-1} = \Delta y_{it} - \alpha \Delta y_{i,t-1}$, see for example Arellano-Bond (1991). Specifying the instrument set as

$$Z_i^D = \begin{bmatrix} y_{i1} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & y_{i1} & \dots & y_{i,T-2} \end{bmatrix}.$$

such that $E[Z_i^{D\prime}\Delta u_i] = 0$ where $\Delta u_i = [\Delta u_{i3}, \Delta u_{i4}, \dots, \Delta u_{iT}]'$, the GMM estimator minimises

$$\left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{D\prime}\Delta u_i\right)' W_N \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{D\prime}\Delta u_i\right)$$

where W_N is a positive semi-definite weight matrix that converges to a positive definite matrix W as $N \to \infty$ (see Hansen (1982)). Under general conditions, an optimal two-step estimator is based on the weight matrix

$$W_N = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^{D\prime} \widehat{\Delta u}_i \widehat{\Delta u}_i^{\prime} Z_i^D \right)^{-1},$$

where $\widehat{\Delta u}_i$ are the residuals based on an initial consistent estimator for α.

Identification For the first-differenced GMM estimator that utilises moment conditions (2.2), the endogenous lagged differences $\Delta y_{i,t-1}$ are instrumented by lagged levels $y_{i1}, ..., y_{i,t-2}$. Clearly, when $\alpha = 1$, the rank condition is not satisfied as $\Delta y_{i,t-1} = v_{i,t-1}$. In this case all these instruments are uncorrelated with the endogenous variable, and therefore α is not identified. Arellano, Hansen and Sentana (1999), henceforth AHS, propose a general test for the identification of the parameters in models estimated by GMM. For the simple AR(1) panel data model considered here, their test of the null hypothesis of underidentification is a test of the validity of the moment conditions

$$E\left(y_i^{t-1} \Delta y_{it}\right) = 0, \tag{2.3}$$

where $y_i^{t-1} = (y_{i1}, ..., y_{i,t-1})'$. When this test rejects, the model is not underidentified. For this model it is clear that a test of underidentification is equivalent to a test of the unit root hypothesis, $H_0 : \alpha = 1$, and we will compare the performance of this AHS test of underidentification to various unit root tests described in the next section. This AHS test is equivalent to the Anderson-Rubin test of $H_0 : \alpha = 1$ in the first-differenced GMM model.

2.1.2. Mean Stationarity, System GMM

If in addition an error components structure is assumed for the error term and mean stationarity of the process is assumed, such that

$$E\left(\eta_i v_{it}\right) = 0 \quad \text{for } i = 1, ..., N \text{ and } t = 2..., T$$

$$y_{i1} = \eta_i + \varepsilon_i \quad \text{for } i = 1, ..., N$$

and

$$E\left(\varepsilon_i\right) = E\left(\eta_i \varepsilon_i\right) = 0 \quad \text{for } i = 1, ..., N,$$

there are the following extra $(T-2)$ linear moment conditions available:

$$E\left(u_{it} \Delta y_{i,t-1}\right) = 0 \quad \text{for } t = 3, ..., T, \tag{2.4}$$

see Arellano-Bover (1995), Ahn-Schmidt (1995) and Blundell-Bond (1998). The so-called system GMM estimator for α is obtained by stacking the residuals from the first-differenced and levels equations, and extending the instrument matrix to

$$Z_i^S = \begin{bmatrix} Z_i^D & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Delta y_{i,T-1} \end{bmatrix}$$

such that $E[Z_i^{S\prime} u_i^+] = 0$ where $u_i^+ = [\Delta u_i', u_{i3}, u_{i4}, \ldots, u_{iT}]'$.

Identification For the system estimator the $T-2$ extra moment conditions (2.4) remain valid when $\alpha = 1$,[2] even though the process is clearly not mean-stationary in this case. Consider the first stage regression for the levels equation, when $T = 3$,

$$y_{i2} = \pi \Delta y_{i2} + r_i.$$

[2] This would not be the case if there were individual-specific drifts.

If $\alpha = 1$, it follows that $\pi = 1$ and $r_i = y_{i1}$. Denote by TP the number of periods that the process has been in existence before the sample is drawn, noting that when $\alpha = 1$, $Var(y_{i1}) \to \infty$ as $TP \to \infty$. Then, for any fixed TP, $\text{plim}_{N\to\infty} \widehat{\pi}_{OLS} = 1$, and the model is (asymptotically, as $N \to \infty$) identified. Therefore, when $Var(y_{i1}) < \infty$, the system GMM estimator can estimate $\alpha = 1$ consistently and thus can also be used to obtain a test for a unit root. For any given sample, the ratio of N to TP determines how well the distribution of the system GMM estimator is then approximated by its (large N) asymptotic distribution.

2.1.3. Covariance Stationarity

For $\alpha < 1$, the processes are covariance stationary when the v_{it} are homoskedastic over time,

$$E\left(v_{it}^2\right) = \sigma_{vi}^2$$

and the initial conditions satisfy

$$y_{i1} = \eta_i + \varepsilon_i$$

with

$$Var\left(\varepsilon_i\right) = \frac{\sigma_{vi}^2}{1-\alpha^2}.$$

In this case there are $(T-2)$ additional linear moment conditions due to the homoskedasticity (through time) of v_{it}, given by

$$E\left(y_{it}u_{it} - y_{i,t-1}u_{i,t-1}\right) = 0 \ \text{ for } \ t = 3, ..., T \tag{2.5}$$

see Ahn-Schmidt (1995). Ahn and Schmidt (1997) further derive the following non-linear moment condition which is valid under the assumption of covariance stationarity

$$E\left[y_{i1}^2+\frac{y_{i2}\Delta u_{i3}}{1-\alpha^2}-\frac{u_{i3}u_{i2}}{(1-\alpha)^2}\right]=0. \tag{2.6}$$

Recently, Kruiniger (2002b) has shown that the non-linear moment condition (2.6) can be replaced by the linear moment condition

$$E\left[(1-\alpha)(\Delta y_{i2})^2+2\Delta y_{i2}\Delta y_{i3}\right]=0. \tag{2.7}$$

The full set of $0.5\times T(T+1)-2$ linear moment conditions under covariance stationarity consists then of (2.2), (2.4), (2.5) and (2.7). The GMM estimator for this model is obtained by again stacking the residuals from the differenced and level equations, augmented by the residual $(\Delta y_{i2})^2+2\Delta y_{i2}\Delta y_{i3}-\alpha(\Delta y_{i2})^2$. The instrument matrix is then given by

$$Z_i^{CS}=\begin{bmatrix} Z_i^D & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & -y_{i2} & 0 & \cdots & \cdots & 0 \\ 0 & \Delta y_{i2} & \cdots & \cdots & 0 & y_{i3} & -y_{i3} & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & & \vdots & 0 & y_{i4} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 & \vdots & \vdots & \ddots & -y_{i,T-1} & \vdots \\ 0 & 0 & \cdots & 0 & \Delta y_{i,T-1} & 0 & 0 & \cdots & y_{iT} & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Identification Kruiniger (2002b) shows that under the null of a unit root, the moment conditions (2.4), (2.5) and (2.7) all remain valid and identify α also when $\alpha=1$ when $Var(y_{i1})<\infty$. However, there is a problem with the estimation of the variance of this GMM estimator when $\alpha=1$ as in that case the information contained in moment condition (2.7) becomes redundant, leading to a singularity in the asymptotic variance.

2.2. Maximum Likelihood

The likelihood of the first-differenced model can be formulated in many different ways, see for example Arellano (2003), Kruiniger (2002a) and Hsiao, Pesaran and Tahmiscioglu (2002). It is the likelihood of the original levels model conditional on the ML estimates of the fixed effects. In the following we adopt the parameterisation of Hsiao et al. (2002). The log-likelihood for the model in first differences under normality is given by

$$\ln L = -\frac{N(T-1)}{2} \ln (2\pi) - \frac{N}{2} \ln |\Omega| - \frac{1}{2} \sum_{i=1}^{N} \Delta v_i^{*\prime} \Omega^{-1} \Delta v_i^{*}, \qquad (2.8)$$

where $\Delta v_i^* = [\Delta y_{i2} - (\alpha - 1)\,\delta_0, \Delta y_{i3} - \alpha \Delta y_{i2}, ..., \Delta y_{iT} - \alpha \Delta y_{i,T-1}]$ and

$$\Omega = \sigma_v^2 \begin{bmatrix} \omega & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \sigma_v^2 \Omega^*$$

with $\omega = \mathrm{Var}\left(\Delta y_{i2}\right) / \sigma_v^2$. This formulation clearly uses homoskedasticity (over individuals and time) and non-serial correlation of the v_{it} explicitly, which could be relaxed in the specification of Ω.

2.2.1. Identification

As shown in the Appendix, the information matrix is singular at $\alpha = 1$ when no further restrictions are imposed on ω (due, for example, to restrictions on the initial conditions). This ML estimator will therefore not identify α when $\alpha = 1$.[3]

[3] Equivalently, the information matrix of the conditonal ML estimator proposed by Lancaster (2002), based on an orthogonal transformation of the fixed effects, is also singular at $\alpha = 1$ in this case because there are no fixed effects, i.e. no individual drifts, when $\alpha = 1$.

2.2.2. Covariance stationarity

Under covariance stationarity, $\omega = 2/(1+\alpha)$. When ω is restricted in this way, the ML estimator also estimates $\alpha = 1$ consistently, and a simple t-test is valid, even though the parameter is on the boundary of the parameter space (see Hsiao et al. (2002) and Kruiniger 2002a).

3. Tests for Unit Roots

3.1. OLS

Under the null $H_0 : \alpha = 1$, the OLS estimator of α in model (2.1) is consistent, and a simple t-test based on this OLS estimator is given by

$$t_{OLS} = \frac{\widehat{\alpha}_{OLS} - 1}{\sqrt{\widehat{\text{Var}}\left(\widehat{\alpha}_{OLS}\right)}}$$

where

$$\widehat{\text{Var}}\left(\widehat{\alpha}_{OLS}\right) = \left(y'_{-1}y_{-1}\right)^{-1}\left(\sum_{i=1}^{N} y'_{i,-1}e_i e'_i y_{i,-1}\right)\left(y'_{-1}y_{-1}\right)^{-1}, \tag{3.1}$$

with $e_i = y_i - y_{i,-1}\widehat{\alpha}_{OLS}$, $y_i = (y_{i2}, ..., y_{iT})'$, $y_{i,-1} = (y_{i1}, ..., y_{i,T-1})'$, and $y_{-1} = \left(y'_{1,-1}, ..., y'_{N,-1}\right)'$. Under the null, t_{OLS} has an asymptotic standard normal distribution as $N \to \infty$ for fixed T. Under the alternative, $\alpha < 1$, the OLS estimator is biased upwards, more so when the variance of η_i is large relative to the variance of v_{it}. The power of this test will therefore depend on the magnitude of $Var(\eta_i)/Var(v_{it})$. Under covariance stationarity and homoskedasticity over individuals, the probability limit of the OLS estimator is given by

$$\text{plim}_{N\to\infty}\,\widehat{\alpha}_{OLS} = \alpha + (1-\alpha)\frac{\sigma_\eta^2}{\sigma_\eta^2 + \frac{\sigma_v^2}{1-\alpha^2}}.$$

3.2. Differencing

In response to this sensitivity to σ_η^2/σ_v^2 of the simple test based on the OLS estimator in the levels equations, Breitung and Meyer (1994) propose a modified Dickey-Fuller statistic, based on the OLS estimator of α in the transformed model

$$y_{it} - y_{i1} = \alpha\left(y_{i,t-1} - y_{i1}\right) + \varepsilon_{it}, \quad t = 3, ..., T, \tag{3.2}$$

where $\varepsilon_{it} = v_{it} - (1-\alpha)\left(y_{i1} - \eta_i\right)$. Clearly, the OLS estimator in this model is also consistent when $\alpha = 1$, in which case

$$\sqrt{N}\left(\widehat{\alpha}_{BM} - 1\right) \rightarrow N\left(0, \sigma_{BM}^2\right)$$

with $\sigma_{BM}^2 = 1/\sum_{j=2}^{T-1}(T-j)$ when the v_{it} are homoskedastic, and a simple t-test is again valid under the null of a unit root. This test would be robust to general forms of heteroskedasticity when constructed using robust standard errors, similar to (3.1). When $\alpha < 1$ this OLS estimator is again upwards biased, however the asymptotic bias does not depend on $Var(\eta_i)/Var(v_{it})$ when the process is mean stationary, and the power of the test is therefore not affected by $Var(\eta_i)/Var(v_{it})$ in that case. Under covariance stationarity and homoskedasticity over individuals, the probability limit of the OLS estimator of α in (3.2) is given by

$$\text{plim}_{N\rightarrow\infty}\,\widehat{\alpha}_{BM} = \frac{\alpha + 1}{2}.$$

We also consider the simple model in first-differences

$$y_{it} - y_{i,t-1} = \alpha\left(y_{i,t-1} - y_{i,t-2}\right) + \left(v_{it} - v_{i,t-1}\right), \quad t = 3, ..., T. \tag{3.3}$$

Under the null of a random walk, the probability limit of the OLS estimator in (3.3) is given by

$$\text{plim}_{N\rightarrow\infty}\,\widehat{\alpha}_{FD} = 1 + \frac{\sum_{i=1}^{N}\sum_{t=3}^{T} v_{i,t-1}\left(v_{it} - v_{i,t-1}\right)}{\sum_{i=1}^{N}\sum_{t=3}^{T} v_{i,t-1}^2} = 0,$$

irrespective of heteroskedasticity of the v_{it}. Therefore, when $\alpha = 1$,

$$\sqrt{N}\left(\widehat{\alpha}_{FD}\right) \rightarrow N\left(0, \sigma_{FD}^2\right)$$

with $\sigma_{FD}^2 = 1/\left(T-2\right)$ if the v_{it} are homoskedastic. Again the variance can easily be estimated allowing for general heteroskedasticity. When $\alpha < 1$ the bias of the estimator is again independent of $Var(\eta_i)/Var(v_{it})$ when the process is mean stationary. Under covariance stationarity and homoskedasticity over individuals, the probability limit of the OLS estimator in (3.3) is given by

$$\text{plim}_{N\rightarrow\infty}\, \widehat{\alpha}_{FD} = \frac{\alpha - 1}{2}.$$

Therefore, the probability limit of the "bias corrected" (under the null) first-differenced OLS estimator $\widehat{\alpha}_{FD} + 1$ is equal to $\left(\alpha + 1\right)/2$, i.e. the same as the probability limit of the OLS estimator in model (3.2).

3.3. Within Groups

Harris and Tzavalis (1999) base a test of the unit root hypothesis on a bias correction of the within groups estimator under the null. Under the assumptions that $v_{it} \sim iid\, N(0, \sigma_v^2)$ and the y_{i1} are fixed observable constants, which implies that y_{i1} is uncorrelated with the sequence $\{v_{it}\}$, Harris and Tzavalis (1999) show that, under the null of a unit root in model (2.1),

$$\sqrt{N}\left(\widehat{\alpha}_{WG} - 1 - P\right) \rightarrow N\left(0, Q\right),$$

where $\widehat{\alpha}_{WG}$ is the within groups estimator of α, and P and Q are given by[4]

$$\begin{aligned} P &= -\frac{3}{T}; \\ Q &= \frac{3\left(17\left(T-1\right)^2 - 20\left(T-1\right) + 17\right)}{5T^3\left(T-2\right)}. \end{aligned}$$

A simple test then is $\left(\widehat{\alpha}_{WG} - 1 - P\right) / \sqrt{Q/N}$, which has an asymptotic standard normal distribution under the null.

As this bias correction and derived variance are valid only under homoskedasticity, it is likely that the test performance will be poor under certain forms of heteroskedasticity. Kruiniger and Tzavalis (2002) extend this approach to allow for general forms of heteroskedasticity and also certain types of serial correlation.

3.4. System GMM

As shown in Section 2.1.2, the system GMM estimator can identify $\alpha = 1$ if the variance of the initial conditions is finite. This estimator is consistent under the null and under mean stationary alternatives. A test for a unit root is then given by the simple t-test, $\left(\widehat{\alpha}_{sys} - 1\right) / se\left(\widehat{\alpha}_{sys}\right)$.

3.5. Maximum Likelihood

Using the likelihood specification (2.8) and imposing the restriction on ω due to covariance stationarity, $\omega = 2/\left(1+\alpha\right)$, results in a consistent estimator under the null and under covariance stationary alternatives. A simple t-test based on this ML estimator can therefore be used to test whether $\alpha = 1$.

[4]Note that these expressions differ from those in Harris and Tzavalis (1999, p.207) due to the fact that our first observation is y_{i1}, not y_{i0}. Therefore, their panel length T is replaced by $T-1$ in our case.

There is a connection between this ML setup and the model of Breitung and Meyer (1994). Setting $\omega = 1$ will result in an ML estimator that is consistent under the null, but biased under the alternative. This ML estimator for α is given by

$$\widehat{\alpha} = \left(\sum_{i=1}^{N} \Delta w_i' \left(\Omega^*\right)^{-1} \Delta w_i \right)^{-1} \sum_{i=1}^{N} \Delta w_i' \left(\Omega^*\right)^{-1} \Delta y_i$$

where $\Delta w_i' = \left(0, \, \Delta y_{i,-1}'\right)$. It is easily seen that this estimator is numerically identical to the OLS estimator in model (3.2) as proposed by Breitung and Meyer (1994). This follows because

$$\begin{bmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i1} \\ \vdots \\ y_{iT} - y_{i1} \end{bmatrix} = R \begin{bmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ \vdots \\ y_{iT} - y_{iT-1} \end{bmatrix}$$

where

$$R = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ & & & & \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

and $R'R = \left(\Omega^*\right)^{-1}$ when $\omega = 1$. Hence the Breitung-Meyer test has an interpretation as a Wald test based on a maximum likelihood estimator that is consistent under the null but not under the alternative.

4. Asymptotic Local Power Comparisons

As shown in the Appendix, the asymptotic local power of these tests depends on whether the processes are covariance stationary or mean stationary under the alternative. The limiting distributions of the t-statistics

$$\sqrt{N} \left(\frac{\widehat{\alpha} - 1}{se\left(\widehat{\alpha}\right)} \right)$$

(or $\sqrt{N}\left(\frac{\widehat{\alpha}}{se(\widehat{\alpha})}\right)$ in the case of the "bias corrected" first-differenced OLS estimator) under local alternatives

$$\alpha = 1 - \frac{c}{\sqrt{N}}$$

are given in Table 1, where the v_{it} are assumed to be homoskedastic over time and individuals. The tests/estimators are denoted OLS, FD, WG, BM, GMM-SYS and MLDCS for the tests based on the levels OLS, first-differenced OLS, within-groups, Breitung-Meyer, system GMM and first-differenced maximum likelihood estimators respectively.[5]

BM and GMM-SYS have the same asymptotic local power, which increases more rapidly with T than WG and FD. The asymptotic local power of GMM-SYS and BM is larger than that of FD and WG for all $T > 3$, whereas the asymptotic local power of WG is larger than that of FD for $T > 5$. For $T = 4$ and $T = 5$ the asymptotic local power of FD and WG are very similar, with that of FD slightly larger in those cases. The asymptotic local power under covariance stationary alternatives is half that under mean stationary alternatives for FD, WG, BM and GMM-SYS. MLDCS has the same asymptotic local power as BM and GMM-SYS under covariance stationary alternatives.

Under mean stationary alternatives, as expected the asymptotic local power of levels OLS depends on the variance parameters in the model. Under covariance stationary alternatives, for the test based on this OLS estimator one has to consider local alternatives of the form

$$\alpha = 1 - \frac{c}{N},$$

[5]MLDCS is the maximum likelihood estimator which imposes the restriction on ω implied by covariance stationarity, giving an estimator that is consistent under the null.

as otherwise the asymptotic variance of the OLS estimator is zero. In that case

$$\sqrt{N}\left(\frac{\widehat{\alpha}_{OLS}-1}{se\left(\widehat{\alpha}_{OLS}\right)}\right)\xrightarrow{d}N\left(-\sqrt{c\frac{(T-1)}{2}},1\right),$$

which is independent of the variance parameters. Clearly, under covariance stationary alternatives, the simple test based on levels OLS has the largest asymptotic local power.

Table 1. Asymptotic distributions of t-test statistics under local alternatives.

	Mean Stationary	Covariance Stationary
OLS*	$N\left(-c\frac{(T-1)\sigma_\varepsilon^2+(T-1)(T-2)\sigma_v^2/2}{\sqrt{\sigma_v^2\left((T-1)\sigma_\eta^2+(T-1)\sigma_\varepsilon^2+(T-1)(T-2)\sigma_v^2/2\right)}},1\right)$	$N\left(-\sqrt{c\frac{(T-1)}{2}},1\right)^*$
FD	$N\left(-c\sqrt{T-2},1\right)$	$N\left(-\frac{c}{2}\sqrt{T-2},1\right)$
WG**	$N\left(-\frac{c}{\sqrt{Q}},1\right)$	$N\left(-\frac{c}{2\sqrt{Q}},1\right)$
BM	$N\left(-c\sqrt{\frac{(T-1)(T-2)}{2}},1\right)$	$N\left(-\frac{c}{2}\sqrt{\frac{(T-1)(T-2)}{2}},1\right)$
GMM-SYS	$N\left(-c\sqrt{\frac{(T-1)(T-2)}{2}},1\right)$	$N\left(-\frac{c}{2}\sqrt{\frac{(T-1)(T-2)}{2}},1\right)$
MLDCS		$N\left(-\frac{c}{2}\sqrt{\frac{(T-1)(T-2)}{2}},1\right)$

* For OLS under covariance stationary alternatives, $\alpha=1-\frac{c}{N}$, in all other cases, $\alpha=1-\frac{c}{\sqrt{N}}$.

** $Q=\frac{3\left(17(T-1)^2-20(T-1)+17\right)}{5T^3(T-2)}$

5. Monte Carlo Results

In this section we present the results of a Monte Carlo study, investigating the properties of the various estimators and test statistics described in the previous sections.

The general data generating process is

$$\begin{aligned} y_{i1} &= \eta_i + \varepsilon_i \\ y_{it} &= \alpha y_{i,t-1} + (1-\alpha)\,\eta_i + v_{it} \end{aligned}$$

with $\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$, $\eta_i \sim N\left(0, \sigma_\eta^2\right)$ and $v_{it} \sim N\left(0, 1\right)$. Under the null we consider random walk processes with different values of σ_ε^2 and hence $Var(y_{i1})$, with higher values corresponding to larger values of TP, or processes that have been generating the data for longer periods prior to the start of our estimation sample. Under the alternative we consider mean-stationary process with unrestricted variances σ_ε^2 and covariance stationary processes with $\sigma_\varepsilon^2 = \frac{\sigma_v^2}{1-\alpha^2}$. For covariance stationary alternatives, we consider different values of σ_η^2, which affects the inconsistency of the levels OLS estimator and the finite sample bias of the GMM estimators. The sample size for all cases is $N = 200$, $T = 6$.

5.1. Estimation Results

Table 2 presents the estimation results for the various estimators, under the null hypothesis that $\alpha = 1$. GMM-DIF denotes the first-differenced GMM estimator, and all GMM results reported are for the efficient two-step estimators. MLD denotes the maximum likelihood estimator in the first-differenced model, not imposing any restrictions on ω.

Table 2. Estimation Results, $\alpha = 1$

		$\sigma_\varepsilon^2 = 50$	$\sigma_\varepsilon^2 = 4$	$\sigma_\varepsilon^2 = 1$
OLS	Mean	0.9999	0.9997	0.9990
	St Dev	0.0044	0.0121	0.0161
WG	Mean	0.4989	0.4993	0.4987
	St Dev	0.0345	0.0347	0.0347
FD	Mean	-0.0001	0.0002	-0.0001
	St Dev	0.0353	0.0356	0.0355
BM	Mean	0.9989	0.9994	0.9989
	St Dev	0.0225	0.0225	0.0225
GMM-DIF	Mean	0.1926	0.1929	0.2012
	St Dev	0.4373	0.4329	0.4306
GMM-SYS	Mean	0.9951	0.9999	0.9997
	St Dev	0.0325	0.0247	0.0245
MLD	Mean	0.9958	0.9958	0.9966
	St Dev	0.0897	0.0902	0.0898
MLDCS	Mean	0.9979	0.9988	0.9979
	St Dev	0.0449	0.0450	0.0449

Notes: based on 10,000 replications. $N = 200$, $T = 6$.

The results for the GMM-DIF, WG, FD, BM, MLD and MLDCS estimators are not affected by the variance of the initial condition. The mean of the GMM-DIF estimates is around 0.2, with a large standard deviation of around 0.43, illustrating the identification problem that we noted in section 2.1.1. The mean of the OLS and BM estimates are both close to one, with OLS having a smaller variance than BM. The mean of the GMM-SYS and MLDCS estimates are both close to one, with GMM-SYS having a smaller variance than MLDCS. However these estimators provide less precise estimates of α under the null than either the BM and, especially, the simple OLS estimator. The mean of the WG estimates is close to 0.5 (the bias $P = -3/T = -0.5$) and the mean of the FD estimates is close to 0, with WG having a slightly smaller variance than FD. The standard deviation of the BM estimator (imposing $\omega = 1$) is half that of the MLCDS estimator (imposing $\omega = 2/(1+\alpha)$). The standard deviation of the OLS estimator decreases with increasing σ_ε^2, whereas the opposite happens with the GMM-SYS estimator. The MLD estimator is centered around 1 with a standard deviation that is approximately twice as large as that of the MLCDS estimator.

Table 3 presents the results for the same estimators under mean stationary alternative hypotheses, for values of $\alpha = \{0.90, 0.95, 0.98\}$, and the initial variance $\sigma_\varepsilon^2 = \{1, 4\}$. MLDCS is inconsistent in this case. The OLS and BM estimators are also inconsistent, although the biases of these estimators diminish as $\alpha \to 1$. Although consistent, GMM-DIF has a downward finite sample bias, that becomes extreme as $\alpha \to 1$. GMM-SYS is virtually unbiased, with more precision when σ_ε^2 is smaller. MLD is virtually unbiased at $\alpha = 0.9$, but has a slight downward bias for larger values of α.

Table 4 presents the estimation results under covariance stationary alternative hypotheses, for the same values of α and for $\sigma_\eta^2 = \{1, 100\}$. Of the consistent estimators, the MLDCS estimator which exploits covariance stationarity now performs best in terms of both bias and precision. The GMM-DIF is again downward biased, much more so when $\sigma_\eta^2 = 100$ than when $\sigma_\eta^2 = 1$. The GMM-SYS and

Table 3. Estimation Results, mean stationary initial conditions

		$\alpha = 0.90$		$\alpha = 0.95$		$\alpha = 0.98$	
		$\sigma_\varepsilon^2 = 4$	$\sigma_\varepsilon^2 = 1$	$\sigma_\varepsilon^2 = 4$	$\sigma_\varepsilon^2 = 1$	$\sigma_\varepsilon^2 = 4$	$\sigma_\varepsilon^2 = 1$
OLS	Mean	0.9179	0.9359	0.9575	0.9629	0.9824	0.9847
	St Dev	0.0136	0.0171	0.0128	0.0166	0.0123	0.0162
WG	Mean	0.4246	0.4040	0.4497	0.4421	0.4762	0.4753
	St Dev	0.0349	0.0349	0.0345	0.0346	0.0349	0.0352
FD	Mean	-0.0589	-0.0727	-0.0378	-0.0441	-0.0181	-0.0186
	St Dev	0.0342	0.0338	0.0341	0.0345	0.0353	0.0351
BM	Mean	0.9383	0.9179	0.9586	0.9515	0.9804	0.9798
	St Dev	0.0237	0.0244	0.0233	0.0235	0.0232	0.0231
GMM-DIF	Mean	0.8321	0.6996	0.7003	0.4729	0.3575	0.2444
	St Dev	0.1370	0.2222	0.2739	0.3727	0.4296	0.4313
GMM-SYS	Mean	0.8914	0.9087	0.9459	0.9492	0.9784	0.9800
	St Dev	0.0465	0.0300	0.0332	0.0272	0.0281	0.0255
MLD	Mean	0.8923	0.8984	0.9341	0.9344	0.9685	0.9676
	St Dev	0.0949	0.0991	0.0934	0.0947	0.0909	0.0907
MLDCS	Mean	0.8781	0.8415	0.9182	0.9047	0.9611	0.9598
	St Dev	0.0459	0.0459	0.0456	0.0458	0.0461	0.0459

Notes: based on 5,000 replications. $N = 200$, $T = 6$.

Table 4. Estimation Results, covariance stationary initial conditions

		$\alpha = 0.90$		$\alpha = 0.95$		$\alpha = 0.98$	
		$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 100$	$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 100$	$\sigma_\eta^2 = 1$	$\sigma_\eta^2 = 100$
OLS	Mean	0.9154	0.9950	0.9542	0.9953	0.9808	0.9960
	St Dev	0.0127	0.0029	0.0095	0.0029	0.0062	0.0028
WG	Mean	0.4355	0.4357	0.4673	0.4675	0.4864	0.4864
	St Dev	0.0349	0.0351	0.0349	0.0348	0.0344	0.0344
FD	Mean	-0.0497	-0.0496	-0.0254	-0.0250	-0.0103	-0.0101
	St Dev	0.0343	0.0342	0.0344	0.0347	0.0350	0.0352
BM	Mean	0.9489	0.9487	0.9740	0.9740	0.9889	0.9784
	St Dev	0.0235	0.0238	0.0229	0.0231	0.0227	0.0451
GMM-DIF	Mean	0.8460	0.5064	0.8504	0.4292	0.7373	0.3860
	St Dev	0.1196	0.3489	0.1665	0.4068	0.2833	0.4305
GMM-SYS	Mean	0.8862	0.9267	0.9313	0.9533	0.9560	0.9693
	St Dev	0.0547	0.0681	0.0595	0.0641	0.0650	0.0628
MLD	Mean	0.8944	0.8960	0.9315	0.9311	0.9628	0.9621
	St Dev	0.0936	0.0932	0.0906	0.0898	0.0891	0.0899
MLDCS	Mean	0.8983	0.8980	0.9482	0.9482	0.9780	0.9784
	St Dev	0.0463	0.0468	0.0455	0.0459	0.0451	0.0451

Notes: based on 10,000 replications. $N = 200$, $T = 6$.

MLD estimators both show some finite sample bias in these experiments.[6] One interesting finding concerns the properties of the simple OLS estimator when the variance ratio σ_η^2/σ_v^2 increases. While this increases the upward bias of the OLS estimator, as expected, we also find that this reduces its variance. The latter may mitigate the effect of the increasing bias on the power of t-tests based on this simple estimator to reject the null hypothesis of $\alpha = 1$.

5.2. Test Results

Table 5 presents the empirical rejection frequencies at a nominal size of 5% for various tests of the null hypothesis that $\alpha = 1$ against the alternative that $\alpha < 1$.[7] UI-DIF denotes the AHS test of underidentification for the GMM-DIF estimator, as described in section 2.1.1. As expected, the t-test based on the MLD estimator has poor size properties, reflecting the fact that the information matrix is singular at $\alpha = 1$. The t-test based on the GMM-SYS estimator rejects too infrequently in the experiment with high σ_ε^2. This is consistent with identification becoming weak for this estimator in the case where the process has been in existence for many periods prior to the estimation sample, as discussed in section 2.1.2. The empirical rejection frequencies are close to the nominal size of 5% for all the other tests considered here, indicating no serious size distortion problems with these tests.

[6] We also considered an extended GMM estimator that exploits the additional moment conditions (2.5) and (2.7) which are valid under covariance stationarity. This estimator had less bias and more precision than GMM-SYS, but did not perform as well as MLDCS (for example, at $\alpha = 0.98$ and $\sigma_\eta^2 = 1$, this estimator had a mean of 0.9733 and a standard deviation of 0.0564). However tests based on this estimator were found to have poor size properties, consistent with the discussion in Kruiniger (2002b), and are not considered here.

[7] The t-test based on the GMM-SYS estimator uses the finite sample corrected variance estimates of Windmeijer (2005). All results are reported for one-sided t-tests.

Table 5. Size properties of tests $H_0 : \alpha = 1$, $H_1 : \alpha < 1$

	$\sigma_\varepsilon^2 = 50$	$\sigma_\varepsilon^2 = 4$	$\sigma_\varepsilon^2 = 1$
OLS	0.0548	0.0555	0.0563
WG	0.0578	0.0557	0.0563
FD	0.0509	0.0523	0.0524
BM	0.0545	0.0550	0.0576
GMM-SYS	0.0271	0.0399	0.0451
MLD	0.0981	0.1041	0.1009
MLDCS	0.0528	0.0539	0.0555
UI-DIF	0.0570	0.0556	0.0539

Notes: based on 10,000 replications. $N = 200$, $T = 6$.

Figures 1 and 2 display the rejection frequencies at the 5% level of the various unit root tests for experiments in which the series are mean stationary with $\alpha = 0.85, 0.86, ..., 0.99, 1$. We report results for two cases with different variances of the initial conditions. In Figure 1 $\sigma_\varepsilon^2 = 4$, whereas in Figure 2 $\sigma_\varepsilon^2 = 1$.[8] In both cases $\sigma_\eta^2 = 1$. In all these experiments, the t-test based on the simple OLS levels estimator is found to have the highest power. The UI-DIF test has high power to reject alternatives with $\alpha < 0.95$ in the experiments with $\sigma_\varepsilon^2 = 4$, but this test has the lowest power to reject mean stationary alternatives that are local to the null. The ranking of the remaining t-tests is in line with the results for asymptotic local power against mean stationary alternatives reported in Table 1. BM has

[8] The properties of the underlying estimators for these cases were considered in Table 3.

more power than WG, which in turn is more powerful than FD. GMM-SYS has the same asymptotic local power as BM, but is affected more by finite sample considerations and displays somewhat less power in both cases. It is interesting to note that MLDCS has exactly the same power as BM in all these experiments, which suggests that there is an exact bias-variance tradeoff.

Figures 3 and 4 display the rejection frequencies at the 5% level of the various unit root tests for experiments in which the series are covariance stationary with $\alpha = 0.85, 0.86, ..., 0.99$. Here we report results for two cases with different variances of the individual effects, and the variance of the initial conditions satisfying covariance stationarity in all cases. In Figure 3 $\sigma_\eta^2 = 1$, whereas in Figure 4 $\sigma_\eta^2 = 100$.[9] For the low value of σ_η^2, the t-test based on the simple OLS levels estimator has the highest power against all the alternatives considered, and the UI-DIF test has notably higher power than any of the other tests. For the high value of σ_η^2, the simple OLS test has the highest power against alternatives that are local to the null of unity, and the UI-DIF test has relatively low power in all these experiments. The t-tests based on the BM and MLDCS again have identical power, and these tests have the highest power to reject the null against covariance stationary values of α below 0.93 in our experiments with $\sigma_\eta^2 = 100$. The ranking of the remaining t-tests is again as suggested by the asymptotic local power calculations reported in Table 1, with BM having more power than WG and FD. The t-test based on the GMM-SYS estimator has similar power to those based on BM and MCDCS in the experiments with low σ_η^2, but has the lowest power of any of these tests against covariance stationary alternatives in the experiments with high σ_η^2.

[9] The properties of the underlying estimators for these cases were considered in Table 4.

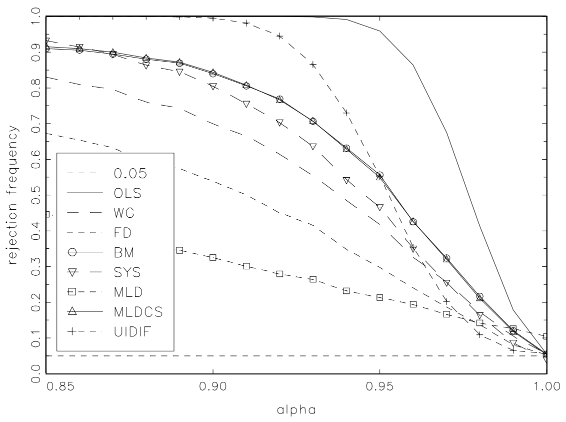

Figure 1. Mean Stationarity, $\sigma_\varepsilon^2 = 4$.

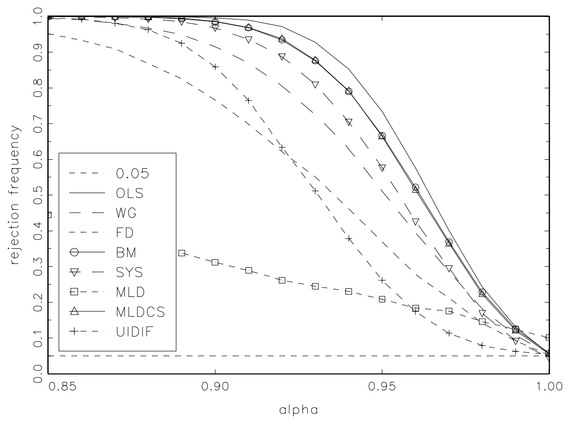

Figure 2. Mean Stationarity, $\sigma_\varepsilon^2 = 1$.

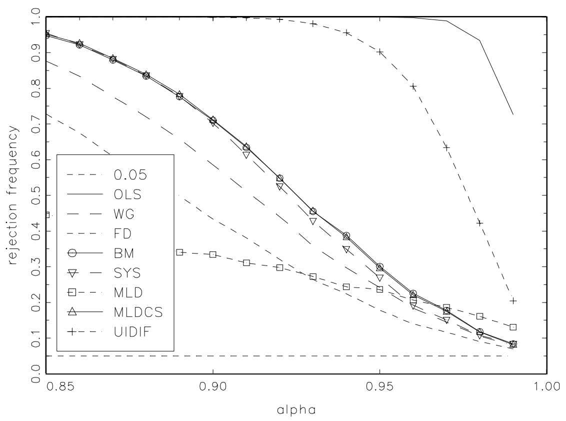

Figure 3. Covariance Stationary, $\sigma_\eta^2 = 1$.

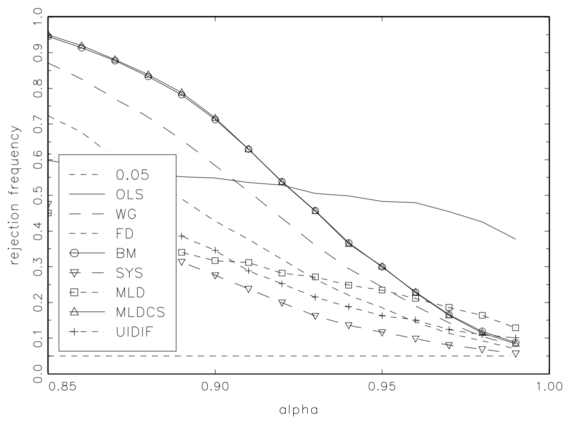

Figure 4. Covariance Stationary, $\sigma_\eta^2 = 100$.

To summarise, the t-test based on the OLS levels estimator is found to perform much better than might have been expected simply on the basis of the sensitivity of the probability limit of this estimator to the relative variance of the error components (i.e. σ_η^2/σ_v^2). This test has the highest power to reject alternatives that are close to the null hypothesis of $\alpha = 1$ in all the cases we consider. For the cases where the simple OLS test does not dominate (i.e. for values of α below 0.93 in Figure 4), the highest power can be obtained using a t-test for the least squares estimator in the transformed model proposed by Breitung and Meyer (1994). Taken together, these findings indicate that these two t-tests based on simple least squares estimators should be considered jointly. Tests based on estimators that are consistent under both the null and under certain alternative hypotheses - such as GMM-SYS and MLDCS - are found to have less power in our experiments than tests based on these least squares estimators that are consistent only under the null.

6. Conclusions

This paper has considered unit root tests in the setting of micro panel data sets with a large cross-section dimension and a small number of time periods. Such tests may correspond to hypotheses of substantive economic interest, or may be studied in order to investigate whether identification based on first-differenced GMM estimators is likely to be weak using the series in question.

We consider a range of unit root tests that have been proposed in this context, providing comparisons based on asymptotic local power calculations and evidence about finite sample properties based on Monte Carlo simulations. Simple t-tests based on least squares estimators that are consistent only under the unit root null are shown to have good size properties and at least as high power as tests based on GMM and ML estimators.

Our results also indicate that rejecting the null hypothesis of a unit root is not sufficient to be confident that consistent GMM estimators will have satisfactory small sample properties. For example, if we consider a mean stationary alternative process with $\alpha = 0.95$ and $\sigma_\varepsilon^2 = \sigma_\eta^2 = 1$, Figure 2 indicates that the t-tests based on least squares estimators will correctly reject the unit root null in around 70% of cases considered. However Table 3 shows that the first-differenced GMM estimator has very poor performance in this case. Similarly for a covariance stationary alternative process with $\alpha = 0.9$ and $\sigma_\eta^2 = 100$, Figure 4 shows that the t-test based on the Breitung-Meyer specification will correctly reject the unit root null in around 70% of cases, while Table 4 shows that the first-differenced GMM estimator also performs poorly in this case. Hence while the poor performance of this GMM estimator in these cases is related to a weak identification problem that becomes extreme in the case of unit root series, our analysis reveals that simply rejecting the unit root null does not establish that first-differenced GMM estimators will provide useful parameter estimates in the same sample.

References

[1] Ahn, S.C. and P. Schmidt, 1995, Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.

[2] Ahn, S.C. and P. Schmidt, 1997, Efficient Estimation of Dynamic Panel Data Models: Alternative Assumptions and Simplified Estimation, *Journal of Econometrics*, 76, 309-321.

[3] Arellano, M., 2003, *Panel Data Econometrics*, Oxford University Press.

[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, 277-298.

[5] Arellano, M. and O. Bover, 1995, Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

[6] Arellano, M., L. Hansen and E. Sentana, 1999, Underidentification?, mimeo, CEMFI, Madrid.

[7] Blundell, R. and S. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.

[8] Breitung, J. and W. Meyer, 1994, Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?, *Applied Economics*, 26, 353-361.

[9] Hall, B.H. and J. Mairesse, 2005, Testing for Unit Roots in Panel Data: An Exploration Using Real and Simulated Data, in D.W.K. Andrews and J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press.

[10] Hansen, L.P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029-1054.

[11] Harris, R.D.F. and E. Tzavalis, 1999, Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed, *Journal of Econometrics*, 91, 201-226.

[12] Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics*, 109, 107-150.

[13] Kruiniger, H., 2002a, Maximum Likelihood Estimation of Dynamic Linear Panel Data Models with Fixed Effects, Working Paper, Queen Mary, University of London.

[14] Kruiniger, H., 2002b, On the Estimation of Panel Regression Models with Fixed Effects, Working Paper, Queen Mary, University of London.

[15] Kruiniger, H. and E. Tzavalis, 2002, Testing for Unit Roots in Short Dynamic Panels with Serially Correlated and heteroskedastic Disturbance Terms, Working Paper, Queen Mary, University of London.

[16] Lancaster, T. (2002) Orthogonal Parameters and Panel Data, *Review of Economic Studies*, 69, 647-666.

[17] Sutton, J. (1997) Gibrat's Legacy, *Journal of Economic Literature*, 35, 40-59.

[18] Windmeijer, F., (2005), A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators, *Journal of Econometrics*, 126, 25-51.

7. Appendix A. Calculations of Asymptotic Distributions of Test Statistics under Local Alternatives

7.1. OLS

7.1.1. Mean Stationarity

$$\begin{aligned} y_{i1} &= \eta_i + \varepsilon_i; \;\; \mathrm{Var}\left(\varepsilon_i\right) = \sigma_\varepsilon^2 \\ y_{it} &= \alpha y_{it-1} + \left(1-\alpha\right)\eta_i + v_{it} \end{aligned}$$

$$\widehat{\alpha} = \frac{\sum_{i=1}^N \sum_{t=2}^T y_{it-1} y_{it}}{\sum_{i=1}^N \sum_{t=2}^T y_{it-1}^2} = \alpha + \frac{\sum_{i=1}^N \sum_{t=2}^T y_{it-1}\left(\left(1-\alpha\right)\eta_i + v_{it}\right)}{\sum_{i=1}^N \sum_{t=2}^T y_{it-1}^2}$$

and so the asymptotic bias b is given by

$$\begin{aligned} b &= \mathrm{plim} \frac{\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T y_{it-1}\left(\left(1-\alpha\right)\eta_i + v_{it}\right)}{\frac{1}{N}\sum_{i=1}^N \sum_{t=2}^T y_{it-1}^2} \\ &= \left(1-\alpha\right)\frac{\left(T-1\right)\sigma_\eta^2}{\left(T-1\right)\sigma_\eta^2 + \sigma_\varepsilon^2 \sum_{j=0}^{T-2}\alpha^{2j} + \sigma_v^2 \sum_{t=3}^T \sum_{j=0}^{t-3}\alpha^{2j}} \end{aligned}$$

The OLS estimator is therefore constently estimating $\alpha + b$. The OLS estimator is unbiased at $\alpha = 1$, and its asymptotic variance is then given by

$$\mathrm{asyvar}\left(\widehat{\alpha}\right)_{\alpha=1} = \frac{\sigma_v^2}{\left(T-1\right)\sigma_\eta^2 + \left(T-1\right)\sigma_\varepsilon^2 + \left(T-1\right)\left(T-2\right)\sigma_v^2/2}$$

For asymptotic local power, consider the sequence

$$\alpha = 1 - \frac{c}{\sqrt{N}}.$$

$$\begin{aligned} \sqrt{N}\left(\frac{\widehat{\alpha}-1}{se\left(\widehat{\alpha}\right)}\right) &= \sqrt{N}\left(\frac{\widehat{\alpha}-\left(\alpha+b\right)}{se\left(\widehat{\alpha}\right)}\right) + \sqrt{N}\left(\frac{\alpha+b-1}{se\left(\widehat{\alpha}\right)}\right) \\ &\rightarrow N\left(0,1\right) - \frac{c - c\frac{\left(T-1\right)\sigma_\eta^2}{\left(T-1\right)\sigma_\eta^2 + \left(T-1\right)\sigma_\varepsilon^2 + \left(T-1\right)\left(T-2\right)\sigma_v^2/2}}{\sqrt{\sigma_v^2/\left(\left(T-1\right)\sigma_\eta^2 + \left(T-1\right)\sigma_\varepsilon^2 + \left(T-1\right)\left(T-2\right)\sigma_v^2/2\right)}} \end{aligned}$$

Therefore the distribution of the t-statistic converges to

$$N\left(-c\frac{(T-1)\,\sigma_{\varepsilon}^{2}+(T-1)\,(T-2)\,\sigma_{v}^{2}/2}{\sqrt{\sigma_{v}^{2}\left((T-1)\,\sigma_{\eta}^{2}+(T-1)\,\sigma_{\varepsilon}^{2}+(T-1)\,(T-2)\,\sigma_{v}^{2}/2\right)}},1\right).$$

Clearly, the power decreases with increasing σ_{η}^{2}.

7.1.2. Covariance Stationarity

When the process is covariance stationary, $\sigma_{\varepsilon}^{2}=\frac{\sigma_{v}^{2}}{1-\alpha^{2}}$, which goes to infinity when $\alpha\rightarrow 1$.The bias b is now given by

$$b=(1-\alpha)\,\frac{\sigma_{\eta}^{2}}{\sigma_{\eta}^{2}+\frac{\sigma_{v}^{2}}{1-\alpha^{2}}}$$

and the asymptotic variance of α is given by

$$\operatorname{asyvar}\left(\widehat{\alpha}\right)=\frac{(1-\alpha)^{2}\,\sigma_{\eta}^{2}\left(1-\frac{\sigma_{\eta}^{2}}{\sigma_{\eta}^{2}+\frac{\sigma_{v}^{2}}{1-\alpha^{2}}}\right)+\sigma_{v}^{2}}{\sigma_{\eta}^{2}+\frac{\sigma_{v}^{2}}{1-\alpha^{2}}}$$

which now converges to zero when $\alpha\rightarrow 1$.

For asymptotic local power therefore consider the sequence

$$\alpha=1-\frac{c}{N}.$$

The leading term in the asymptotic variance is

$$\begin{aligned}\frac{\sigma_{v}^{2}}{(T-1)\left(\sigma_{\eta}^{2}+\frac{\sigma_{v}^{2}}{1-\alpha^{2}}\right)} &= \frac{\left(1-\alpha^{2}\right)\sigma_{v}^{2}}{(T-1)\left(\left(1-\alpha^{2}\right)\sigma_{\eta}^{2}+\sigma_{v}^{2}\right)}\\ &= \frac{\left(\frac{2c}{N}-\frac{c^{2}}{N^{2}}\right)\sigma_{v}^{2}}{(T-1)\left(\left(\frac{2c}{N}-\frac{c^{2}}{N^{2}}\right)\sigma_{\eta}^{2}+\sigma_{v}^{2}\right)}\end{aligned}$$

and so N asyvar $(\widehat{\alpha})$ converges to $2c/(T-1)$.

$$\begin{aligned}\sqrt{N}\left(\frac{\widehat{\alpha}-1}{se(\widehat{\alpha})}\right) &= \sqrt{N}\left(\frac{\widehat{\alpha}-(\alpha+b)}{se(\widehat{\alpha})}\right)+\sqrt{N}\left(\frac{\alpha+b-1}{se(\widehat{\alpha})}\right) \\ &\rightarrow N(0,1)-\frac{c}{\sqrt{2c/(T-1)}}\end{aligned}$$

Therefore the distribution of the t-statistic converges to

$$N\left(-\sqrt{c\frac{(T-1)}{2}},1\right).$$

which is now independent of the two variance components.

7.2. First differenced model

7.2.1. Mean Stationarity

$$\Delta y_{it}=\alpha\Delta y_{it-1}+\Delta v_{it}$$

$$\widehat{\alpha}=\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}\Delta y_{it}}{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}^{2}}=\alpha+\frac{\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}\Delta v_{it}}{\sum_{i=1}^{N}\Delta y_{it-1}^{2}}$$

The asymptotic bias is given by

$$\begin{aligned}b &= \operatorname{plim}\frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}\Delta v_{it}}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}^{2}} \\ &= -\frac{(T-2)\sigma_v^2}{(T-2)\sigma_v^2+(1-\alpha)^2\sigma_v^2\sum_{t=4}^{T}\sum_{j=0}^{t-4}\alpha^{2j}+(1-\alpha)^2\sigma_\varepsilon^2\sum_{j=0}^{T-3}\alpha^{2j}}\end{aligned}$$

The estimator is $\widehat{\alpha}$ is therefore consistently estimating $\alpha+b$. At $\alpha=1$, $b=-1$ and therefore a consistent estimator at $\alpha=1$ is given by $\widehat{\alpha}+1$, with asymptotic variance when $\alpha=1$ given by

$$\text{asyvar}(\widehat{\alpha})_{\alpha=1}=\frac{1}{T-2}$$

For asymptotic local power, consider the sequence

$$\alpha = 1 - \frac{c}{\sqrt{N}}.$$

$$\begin{aligned}
\sqrt{N}\left(\frac{(\widehat{\alpha}+1)-1}{se\left(\widehat{\alpha}\right)}\right) &= \sqrt{N}\left(\frac{\widehat{\alpha}-(\alpha+b)}{se\left(\widehat{\alpha}\right)}\right)+\sqrt{N}\left(\frac{\alpha+b+1-1}{se\left(\widehat{\alpha}\right)}\right) \\
&\rightarrow N(0,1) - c\sqrt{T-2}
\end{aligned}$$

Therefore the distribution of the t-statistic converges to

$$N\left(-c\sqrt{T-2}, 1\right).$$

7.2.2. Covariance Stationarity

$$b = \text{plim}\,\frac{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}\Delta v_{it}}{\frac{1}{N}\sum_{i=1}^{N}\sum_{t=3}^{T}\Delta y_{it-1}^{2}} = -\frac{\sigma_v^2}{2(1-\alpha)\frac{\sigma_v^2}{1-\alpha^2}} = -\frac{1+\alpha}{2}$$

For asymptotic local power, consider the sequence

$$\alpha = 1 - \frac{c}{\sqrt{N}}.$$

$$\begin{aligned}
\sqrt{N}\left(\frac{(\widehat{\alpha}+1)-1}{se\left(\widehat{\alpha}\right)}\right) &= \sqrt{N}\left(\frac{\widehat{\alpha}-(\alpha+b)}{se\left(\widehat{\alpha}\right)}\right)+\sqrt{N}\left(\frac{\alpha+b+1-1}{se\left(\widehat{\alpha}\right)}\right) \\
&\rightarrow N(0,1) - \left(c-\frac{c}{2}\right)\sqrt{T-2}
\end{aligned}$$

Therefore the distribution of the t-statistic converges to

$$N\left(-\frac{c}{2}\sqrt{T-2}, 1\right).$$

and so the power is less under covariance stationarity.

7.3. Breitung-Meyer

7.3.1. Mean Stationarity

$$\begin{aligned} y_{it} - y_{i1} &= \alpha\left(y_{i,t-1} - y_{i1}\right) + (1-\alpha)\,\eta_i + v_{it} - (1-\alpha)\,y_{i1} \\ &= \alpha\left(y_{i,t-1} - y_{i1}\right) + v_{it} - (1-\alpha)\,\varepsilon_i \end{aligned}$$

$$\widehat{\alpha} = \alpha + \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}\left(y_{i,t-1} - y_{i1}\right)\left(v_{it} - (1-\alpha)\,\varepsilon_i\right)}{\sum_{i=1}^{N}\sum_{t=3}^{T}\left(y_{i,t-1} - y_{i1}\right)^2}$$

Asymptotic bias:

$$b = (1-\alpha)\frac{(T-2)\sigma_\varepsilon^2 - \sigma_\varepsilon^2\sum_{t=3}^{T}\alpha^{t-2}}{(T-2)\sigma_\varepsilon^2 - 2\sigma_\varepsilon^2\sum_{t=3}^{T}\alpha^{t-2} + \sigma_\varepsilon^2\sum_{t=3}^{T}\alpha^{2(t-2)} + \sigma_v^2\sum_{t=3}^{T}\sum_{j=0}^{t-3}\alpha^{2j}}$$

Asymptotic variance when there is a unit root:

$$\text{asyvar}\left(\widehat{\alpha}\right)_{\alpha=1} = \frac{\sigma_v^2}{\sum_{t=3}^{T}\sum_{j=1}^{t-2}\sigma_v^2} = \frac{2}{(T-1)\,(T-2)}$$

Therefore,

$$\sqrt{N}\left(\frac{\widehat{\alpha}-1}{se\left(\widehat{\alpha}\right)}\right) \xrightarrow{d} N\left(-c\sqrt{\frac{(T-1)\,(T-2)}{2}}, 1\right)$$

when $\alpha = 1 - \frac{c}{\sqrt{N}}$.

7.3.2. Covariance stationarity

Asymptotic bias

$$b = \frac{1-\alpha}{2}$$

and so

$$\sqrt{N}\left(\frac{\widehat{\alpha}-1}{se\left(\widehat{\alpha}\right)}\right) \xrightarrow{d} N\left(-\frac{c}{2}\sqrt{\frac{(T-1)\,(T-2)}{2}}, 1\right)$$

when $\alpha = 1 - \frac{c}{\sqrt{N}}$.

7.4. System Estimator

7.4.1. Mean Stationarity

$T = 3$. The system GMM estimator is consistent. The limiting variance of the efficient two-step estimator is given by

$$\sigma^2 = \left(E\left(x'Z\right)\left(E\left(Z'uu'Z\right)\right)^{-1} E\left(Z'x\right)\right)^{-1}$$

where

$$\begin{aligned} x &= \begin{bmatrix} \Delta y_2 \\ y_2 \end{bmatrix} \\ Z &= \begin{bmatrix} y_1 & 0 \\ 0 & \Delta y_2 \end{bmatrix} \\ u &= \begin{bmatrix} \Delta v_3 \\ (1-\alpha)\,\eta + v_3 \end{bmatrix} \end{aligned}$$

$$E\left(Z'uu'Z\right) = \begin{bmatrix} 2\sigma_v^2\left(\sigma_\eta^2 + \sigma_\varepsilon^2\right) & -\left(1-\alpha\right)\sigma_v^2\left(\sigma_\eta^2+\sigma_\varepsilon^2\right) \\ -\left(1-\alpha\right)\sigma_v^2\left(\sigma_\eta^2+\sigma_\varepsilon^2\right) & \left(\left(1-\alpha\right)^2\sigma_\eta^2+\sigma_v^2\right)\left(\left(1-\alpha\right)^2\sigma_\varepsilon^2+\sigma_v^2\right) \end{bmatrix}$$

Further,

$$E\left(Z'x\right) = \begin{bmatrix} -\left(1-\alpha\right)\sigma_\varepsilon^2 \\ \sigma_v^2 - \alpha\left(1-\alpha\right)\sigma_\varepsilon^2 \end{bmatrix}$$

when $\alpha \to 1$, σ^2 approaches 1, and so when

$$\alpha = 1 - \frac{c}{\sqrt{N}}$$

$$\begin{aligned} \sqrt{N}\left(\frac{\widehat{\alpha}-1}{se\left(\widehat{\alpha}\right)}\right) &= \sqrt{N}\left(\frac{\widehat{\alpha}-\alpha}{se\left(\widehat{\alpha}\right)}\right) + \sqrt{N}\left(\frac{\alpha-1}{se\left(\widehat{\alpha}\right)}\right) \\ &\to N\left(0,1\right) - c \end{aligned}$$

Therefore the distribution of the t-statistic converges to

$$N\left(-c, 1\right)$$

when $\alpha = 1 - \frac{c}{\sqrt{N}}$.

7.4.2. Covariance Stationarity

When $\sigma_\varepsilon^2 = \frac{\sigma_v^2}{1-\alpha^2}$,

$$E\left(Z'uu'Z\right) = \begin{bmatrix} 2\sigma_v^2\left(\sigma_\eta^2 + \frac{\sigma_v^2}{1-\alpha^2}\right) & -\left(1-\alpha\right)\sigma_v^2\sigma_\eta^2 - \sigma_v^2\frac{\sigma_v^2}{1+\alpha} \\ -\left(1-\alpha\right)\sigma_v^2\sigma_\eta^2 - \sigma_v^2\frac{\sigma_v^2}{1+\alpha} & \left(\left(1-\alpha\right)^2\sigma_\eta^2 + \sigma_v^2\right)\left(\left(1-\alpha\right)\frac{\sigma_v^2}{1+\alpha} + \sigma_v^2\right) \end{bmatrix}.$$

Further,

$$E\left(Z'x\right) = \begin{bmatrix} -\frac{\sigma_v^2}{1+\alpha} \\ \frac{\sigma_v^2}{1+\alpha} \end{bmatrix}$$

In this case $\sigma^2 \to 4$ when $\alpha \to 1$ and so the distribution of the t-statistic converges to

$$N\left(-\frac{c}{2}, 1\right)$$

when $\alpha = 1 - \frac{c}{\sqrt{N}}$.

For general T, the variance of the system estimator decreases at rate $\frac{(T-1)(T-2)}{2}$ and therefore the asymptotic local power of the system estimator is the same as for the Breitung-Meyer test.

7.5. MLDCS

Kruiniger (2002) derives the asymptotic variance of the MLDCS estimator. The asymptotic variance of the ML estimator is 4 when $\alpha = 1$ and $T = 3$. The rate of decrease in the variance is $\frac{(T-1)(T-2)}{2}$, and so for the MLDCS estimator when the process is covariance stationary:

$$\sqrt{N}\left(\frac{\widehat{\alpha}-1}{se\left(\widehat{\alpha}\right)}\right) \xrightarrow{d} N\left(-\frac{c}{2}\sqrt{\frac{\left(T-1\right)\left(T-2\right)}{2}}, 1\right).$$

8. Appendix B. Derivation of Singularity of Information Matrix for MLD when $\alpha = 1$

The model is

$$\begin{aligned} y_{i1} &= \delta_0 + \delta_1 \eta_i + \varepsilon_i \\ y_{it} &= \alpha y_{i,t-1} + (1-\alpha)\eta_i + v_{it} \end{aligned}$$

For the ML estimator as presented by Pesaran et al. (2002) the log-likelihood for the model in differences is given by

$$\ln L = -\frac{N(T-1)}{2}\ln(2\pi) - \frac{N}{2}\ln|\Omega| - \frac{1}{2}\sum_{i=1}^{N}\Delta v_i^{*\prime}\Omega^{-1}\Delta v_i^{*},$$

where $\Delta v_i^* = [\Delta y_{i2} - c^*, \Delta y_{i3} - \alpha\Delta y_{i2}, ..., \Delta y_{iT} - \alpha\Delta y_{i,T-1}]$, $c^* = (\alpha - 1)\delta_0$, and

$$\Omega = \sigma_v^2 \begin{bmatrix} \omega & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \sigma_v^2 \Omega^*$$

with $\omega = \mathrm{Var}(\Delta y_{i2})/\sigma_v^2$. The first-order derivatives of the likelihood functions result in (see Pesaran et al.(2002))

$$\begin{aligned} \widehat{\theta} &= \begin{pmatrix} \widehat{c}^* \\ \widehat{\alpha} \end{pmatrix} = \left(\sum_{i=1}^{N}\Delta W_i'\left(\widehat{\Omega}^*\right)^{-1}\Delta W_i\right)^{-1}\sum_{i=1}^{N}\Delta W_i'\left(\widehat{\Omega}^*\right)^{-1}\Delta y_i \\ \widehat{\sigma}_v^2 &= \frac{1}{N(T-1)}\sum_{i=1}^{N}\left(\Delta y_i - \Delta W_i\widehat{\theta}\right)'\left(\widehat{\Omega}^*\right)^{-1}\left(\Delta y_i - \Delta W_i\widehat{\theta}\right) \\ \widehat{\omega} &= \frac{(T-2)}{T-1} + \frac{1}{\widehat{\sigma}_v^2 N(T-1)^2}\sum_{i=1}^{N}\left[\left(\Delta y_i - \Delta W_i\widehat{\theta}\right)'\Psi\left(\Delta y_i - \Delta W_i\widehat{\theta}\right)\right], \end{aligned}$$

where

$$\Delta W_i = \begin{bmatrix} 1 & 0 \\ 0 & \Delta y_{i2} \\ 0 & \Delta y_{i3} \\ \vdots & \vdots \\ 0 & \Delta y_{i,T-1} \end{bmatrix}$$

and

$$\Psi = \begin{bmatrix} (T-1)^2 & (T-1)(T-2) & (T-1)(T-2) & \dots & T-1 \\ (T-1)(T-2) & (T-2)^2 & (T-2)(T-3) & \dots & T-2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ T-1 & (T-2) & (T-3) & \dots & 1 \end{bmatrix}.$$

The second-derivatives of the log-likelihood function are given by

$$\begin{aligned}
\frac{\partial^2 \ln L}{\partial\theta\partial\theta'} &= -\frac{1}{\sigma_v^2}\sum_{i=1}^{N}\Delta W_i'\left(\Omega^*\right)^{-1}\Delta W_i \\
\frac{\partial^2 \ln L}{\partial\theta\partial\sigma_v^2} &= -\frac{1}{\sigma_v^4}\sum_{i=1}^{N}\Delta W_i'\left(\Omega^*\right)^{-1}\left(\Delta y_i - \Delta W_i\theta\right) \\
\frac{\partial^2 \ln L}{\partial\theta\partial\omega} &= -\frac{1}{\sigma_v^2\left[1+(T-1)(\omega-1)\right]^2}\sum_{i=1}^{N}\Delta W_i'\Psi\left(\Delta y_i - \Delta W_i\theta\right) \\
\frac{\partial^2 \ln L}{\partial\left(\sigma_v^2\right)^2} &= \frac{N(T-1)}{2\sigma_v^4} - \frac{1}{\sigma_v^6}\sum_{i=1}^{N}\left(\Delta y_i - \Delta W_i\theta\right)'\left(\Omega^*\right)^{-1}\left(\Delta y_i - \Delta W_i\theta\right) \\
\frac{\partial^2 \ln L}{\partial\sigma_v^2\partial\omega} &= -\frac{1}{2\sigma_v^4\left[1+(T-1)(\omega-1)\right]^2}\sum_{i=1}^{N}\left(\Delta y_i - \Delta W_i\theta\right)'\Psi\left(\Delta y_i - \Delta W_i\theta\right) \\
\frac{\partial^2 \ln L}{\partial\omega^2} &= \frac{N(T-1)^2}{2\left[1+(T-1)(\omega-1)\right]^2} \\
&\quad -\frac{T-1}{\sigma_v^2\left[1+(T-1)(\omega-1)\right]^3}\sum_{i=1}^{N}\left(\Delta y_i - \Delta W_i\theta\right)'\Psi\left(\Delta y_i - \Delta W_i\theta\right)
\end{aligned}$$

When $\alpha = 1$, and therefore $\omega = 1$ and $c^* = 0$, the information matrix is singular. As in this case

$$
\begin{aligned}
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\theta\partial\theta'} &= -\begin{bmatrix} \frac{T-1}{\sigma_v^2} & 0 \\ 0 & \frac{(T-1)(T-2)}{2} \end{bmatrix} \\
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\theta\partial\sigma_v^2} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\theta\partial\omega} &= -\begin{bmatrix} 0 \\ \frac{(T-1)(T-2)}{2} \end{bmatrix} \\
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\left(\sigma_v^2\right)^2} &= -\frac{T-1}{2\sigma_v^4} \\
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\sigma_v^2\partial\omega} &= -\frac{T-1}{2\sigma_v^2} \\
E\frac{1}{N}\frac{\partial^2 \ln L}{\partial\omega^2} &= -\frac{\left(T-1\right)^2}{2}
\end{aligned}
$$

the information matrix is given by

$$
I\left(\theta,\sigma_v^2,\omega\right) = \begin{bmatrix} \left(T-1\right)/\sigma_v^2 & 0 & 0 & 0 \\ 0 & \left(T-1\right)\left(T-2\right)/2 & 0 & \left(T-1\right)\left(T-2\right)/2 \\ 0 & 0 & \left(T-1\right)/2\sigma_v^4 & \left(T-1\right)/2\sigma_v^2 \\ 0 & \left(T-1\right)\left(T-2\right)/2 & \left(T-1\right)/2\sigma_v^2 & \left(T-1\right)^2/2 \end{bmatrix}
$$

which is clearly singular.