Mean Group Tests for Stationarity in Heterogenous Panels*

Yongcheol Shin
Department of Economics, University of Edinburgh

Andy Snell
Department of Economics, University of Edinburgh

This version, August 2002

Abstract

This paper proposes the panel-based mean group tests for the null of stationarity against the alternative of unit roots in the presence of both heterogeneity across cross-section units and serial correlation across time periods. Using both sequential and joint asymptotic analyses the proposed test statistic is shown to be distributed as standard normal under the null for large N (number of groups) and large T (number of time periods).

Monte Carlo results support the use of join asymptotic limits (under further condition that $N/T \to 0$) as a guide to finite sample performance, but also clearly indicate that the power of our suggested panel-based test is substantially higher than that of the single time series-based test.

JEL Classification: C12, C15, C22, C23.
Key Words: Mean Group Tests, Heterogeneous Panels, Joint Asymptotic Theory, Stationarity, Unit Roots, Monte Carlo Simulation, Finite Sample Adjustment.

*We are grateful to Kaddour Hadri, George Kapetanios, Peter Phillips and Ron Smith for helpful comments. We also gratefully acknowledge partial financial support from the ESRC (grant No. R000223399). The usual disclaimer applies.

1 Introduction

In recent years there has been an upsurge in the availability and use of panel data sets where both the number of cross sectional panel units (N) and the number of time series observations (T) are very large. For example Summers and Heston's (1991) large multi-country panel data has been and still is the focus of much empirical work in the area of macroeconomic growth. Furthermore, as time progresses, micro panel data sets such as the British Household Panel Survey (BHPS), where T is not currently so large, are being updated to incorporate new time series observations as they become available. It is now recognised that when used in an appropriately rigorous fashion large panels can be hugely informative about the unknown parameters of economic models and yield very powerful tests of hypotheses nested within these models.

Unlike their small T-large N counterparts, the large N-large T panels create new econometric challenges not only to develop new estimators and test statistics but also to solve the technical difficulties raised in asymptotic analysis where both N and T go to infinity. There now exists a substantial amount of literature that extends traditional single index (either large T or large N but not both) asymptotic theory to the double indexed case (large N and large T together). One main research area for analysing these large T and large N panels has been an extension of the single time series-based unit root or cointegration tests into the panel-based ones. See for example, Levin and Lin (1993), Quah (1994), McCoskey and Kao (1998), Maddala and Wu (1999), Pedroni (1999), Kao (1999), Hadri (2000), Choi (2001) and Im *et al.* (2002). See also the survey papers by Banerjee (1999), Smith (2000) and Baltagi and Kao (2002) for related issues.

Nearly all of the papers in this burgeoning literature adopt approximations based on sequential asymptotic theory, which assumes that the time (T) and cross-section (N) dimensions grow infinitely large in strict sequence, namely T first followed by N. However, from a practitioner's viewpoint, sequential asymptotic theory seems somewhat artificial because one is dealing with data where T and N are large *together*. An exception is Phillips and Moon (1999) who provide rigorous joint asymptotic analysis of pooled estimators obtained from (static) regressions in panels with nonstationary regressors when the underlying regression disturbances follow general (linear) stationary processes. In this analysis, if the additional condition that $N/T \to 0$ holds, then they show that sequential asymptotic results for their pooled estimators would be equivalent to the joint ones. See also Kauppi (2000) for a joint asymptotic analysis of pooled estimates in the context of panels containing near integrated regressors with heterogeneous localising parameters. These exceptions apart, joint asymptotic results have rarely been used for inference in studies to date. The key reason why sequential asymptotic analysis continues to be used lies almost certainly in its simplicity. In fact, this type of analysis is usually little different from its conventional single index (single equation-large T) counterpart and thus the underlying statistical assumptions used in conventional single equation analysis rarely need to be altered.

Another important modelling problem in large T-large N panels is the extent of cross-sectional heterogeneity. This may be so large as to preclude the use of pooling [the method adopted by Phillips and Moon (1999) and Kauppi (2000)]. Since it is now possible to estimate a separate regression for each panel unit (which is not possible in the small T case), it is also natural to think of heterogeneous panels where the parameters can differ over cross-section units. An approach that is becoming increasingly popular in this context is to focus estimation and inference on so called mean group quantities that are "averages" across panel units. This approach has been advanced by Pesaran and Smith (1995) for estimation, and first applied to the panel-based unit root test by Im *et al.* (2002).

Our test is based on the mean of the KPSS stationarity test statistics from each panel unit [see Kwiatokwski *et al.* (1992)], which in turn are computed using parametric estimation of the long-run variance of the underlying serially correlated disturbances. Using both *sequential* and *joint* asymptotic theory where T and N are allowed to grow large in sequence and together, respectively, we show that the suggested test statistic has a standard normal distribution as both N and T grow without bounds under the null hypothesis of stationarity. More important perhaps is that the joint asymptotic approach also predicts that unless N/T is small, then the asymptotic limit will fail, which suggests that sequential asymptotic analysis may be misleading as a guide to using mean group tests in small samples.

The finite sample performances of the proposed mean group statistic are examined using Monte Carlo experiments. The simulation results seem to support the joint asymptotic theory quite well in the sense that the size of the tests is close to the nominal level while retaining significantly higher power than obtained in the single time series case. But, there are some situations where the suggested test tends to over-reject. This over-size problem, which is of more concern for the stationarity test in the presence of serial correlation, is shown to be greatly reduced by a finite-sample adjustment based on the non-parametric group-based estimator of the individual stationarity statistic, which is obtained empirically. In particular, this modification seems to be very effective in pseudo panels where $T \geq N$. Another practically important finding is that the mean group stationarity tests, when constructed in conjunction with the *non-parametric* long-run variance estimator suggested by KPSS, is of no use with the typical sample sizes encountered because of massive size distortions in almost all cases considered.

The layout of the paper is as follows. Section 2 discusses underlying models and assumption. Section 3 presents the mean group KPSS stationarity test. Section 4 examines the small sample performance of the suggested statistics. Section 5 provides some concluding remarks. All the proofs are stored in the Mathematical Appendix.

2 Model

We suppose that the stochastic process, y_{it}, consists of unobserved components over a sample of N cross-sections and T time periods:

$$y_{it} = \alpha_i + e_{it},\ i = 1, ..., N;\ t = 1, ...T, \tag{2.1}$$

where

$$e_{it} = \gamma_{it} + u_{it};\ \gamma_{it} = \gamma_{it-1} + v_{it}, \tag{2.2}$$

α_i's are unknown parameters. Since the intercepts are included in each individual regressions, it is assumed that $\gamma_{i0} = 0$ for all i without loss of generality. The above model can be easily extended to allow for linear time trends with heterogenous coefficients. We assume that the u_{it} for $i = 1, ..., N$ and $t = 1, ..., T$, are independently distributed stationary variates with zero means and finite (possibly heterogenous) variances, $\sigma_{u_i}^2$, and that the v_{it} are *iid* variates with zero means and variances, $\sigma_{v_i}^2$.

We consider testing the null hypothesis that all y_{it}'s are stationary (around deterministic components),

$$H_0 : \sigma_{v_1}^2 = \cdots = \sigma_{v_N}^2 = 0, \tag{2.3}$$

against the alternatives

$$H_1 : \sigma_{v_i}^2 > 0,\ i = 1, ..., N_1;\ \sigma_{v_i}^2 = 0,\ i = N_1 + 1, ..., N_2. \tag{2.4}$$

This alternative hypothesis allows for $\sigma_{v_i}^2$ to differ across groups and includes the homogeneous alternative of $\sigma_{v_i}^2 = \sigma_v^2 > 0$ for all i as a special case. It also implies that some of the individual series may be stationary under the alternative. As will be shown below, the consistency of the proposed panel stationarity test is ensured if the fraction of the individual processes that are unit root is non-zero under (2.4).

3 Testing for Stationarity in Heterogeneous Panels

In this section we will use the 'mean group test' approach advanced by Pesaran and Smith (1995) and Im *et al.* (2002) in the context of estimating dynamic heterogeneous panels, where they showed that the conventional pooled estimator is inconsistent in such a situation. While a test based on pooled estimates can also be employed, it might result in misleading inferences in panels with heterogeneous dynamics, as shown via Monte Carlo simulation studies in Im *et al.* (2002). By contrast, the mean group approach takes full account of heterogeneity in panel units and is thus a more natural vehicle for testing in such contexts. It yields a test that is consistent against several types of partial departures from the null and exploits the panel dimension of the data without having to introduce the homogeneity assumption that would allow pooled estimation.

In this section, we utilise both joint and sequential asymptotic theories in deriving limits of the mean group test statistics and analyse their role in guiding empirical practice.

3.1 The Case with Serially Uncorrelated Errors

We begin with a simple case where the underlying stationary disturbances u_{it} are Gaussian white noise, and assume:

Assumption 3.1 *The u_{it}'s in (2.2) are independent normal variates with zero means and finite heterogenous variances, $\sigma_{u_i}^2 > 0$.*

The statistic for stationarity for the individual group is defined by (see Kwiatkowski *et al.*, 1992, hereafter KPSS)

$$\eta_{iT} = \frac{T^{-2} \sum_{t=1}^{T} s_{it}^2}{\hat{\sigma}_{iT}^2}, \tag{3.1}$$

where $s_{it} = \sum_{j=1}^{t} \hat{e}_{ij}$ is the partial sum process of the OLS residual obtained from the regression of y_{it} on constant (*i.e.* demeaned data), and

$$\hat{\sigma}_{iT}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{it}^2. \quad (3.2)$$

Under the null hypothesis of stationarity, KPSS have shown that as $T \to \infty$,

$$\eta_{iT} \overset{T}{\Rightarrow} \eta_i = \int_0^1 V(r)^2 dr, \quad (3.3)$$

where $V(r) = W(r) - rW(1)$ is a standard Brownian bridge, and $W(r)$ is a standard Brownian motion defined on $r \in [0,1]$, and where here and henceforth, $\overset{T}{\Rightarrow} \left(\overset{N}{\Rightarrow}\right)$denotes weak convergence as $T \to \infty$ $(N \to \infty)$. Under the alternative hypothesis of a unit root, η_{iT} diverges to infinity. The mean and variance of the standard Brownian bridge are

$$\mu = E\left(\int_0^1 V(r)dr\right) = \frac{1}{6}; \; \omega^2 = Var\left(\int_0^1 V(r)dr\right) = \frac{1}{45}. \quad (3.4)$$

See, for example, Hadri (2000) for an analytic derivation via the characteristic function method.

In this paper we consider the following panel-based statistic for stationarity:

$$\tau_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\eta_{iT} - \mu}{\omega}\right). \quad (3.5)$$

Under very general assumptions for the u_{it}'s including those of Assumption 3.1 above as a special case, it is easy to show that this statistic weakly converges to a standard normal variate sequentially with $T \to \infty$, followed by $N \to \infty$.

More specifically, for each i, as $T \to \infty$, the individual statistic η_{iT} weakly converges to η_i of (3.3), so that we may write

$$\tau_{NT} \overset{T}{\Rightarrow} \tau_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\eta_i - \mu}{\omega}\right). \quad (3.6)$$

Next, using the fact that the η_i's are *iid* with mean μ and variance ω^2 under Assumption 3.1, and by invoking the Lindberg-Levy central limit theorem (CLT), we have as $N \to \infty$,

$$\tau_N \overset{N}{\Rightarrow} \tau \sim N(0,1). \quad (3.7)$$

In sum, the prediction of sequential asymptotic theory is that the mean group statistic τ_{NT} will be asymptotically standard normal. However, this asymptotic result may be misleading in small samples empirically unless N/T is small, as the following joint asymptotic analysis will show. In particular, the following theorem derives the condition on N and T for the standard normal limit to be reliable in small samples:

Theorem 3.1 *Under Assumption 3.1, and under the null hypothesis (2.3), as* $T \to \infty$ *and* $N \to \infty$ *with* $N/T \to 0$*, the* τ_{NT} *statistic defined by (3.5) weakly converges to a standard normal variate. Under the alternative hypothesis (2.4), as* $T, N \to \infty$ *and* $N_1/N \to \delta > 0$*,* τ_{NT} *diverges to infinity.*

Proof. See the Appendix. ■

In the case where y_{it} is regressed on an intercept and a linear time trend, it has also been shown (*e.g.* KPSS) that the individual stationarity test statistic converges in distribution to the random variable,

$$\eta_{iT} \overset{T}{\Rightarrow} \eta_i = \int_0^1 V_2(r)^2 dr$$

where $V_2(r) = [W(r) + (2r - 3r^2) W(1) + 6r(r-1) \int_0^1 W(s)\, ds$ is a second level Brownian Bridge process. In this case the construction of the panel stationarity test statistic would be slightly modified, *i.e.* we replace the de-meaned process by the de-trended one, and use the mean and variance of the integral of the second-level standard Brownian Bridge process which are given by [*e.g.* Hadri (2000)]

$$\mu = E\left[\int_0^1 V_2(r)dr\right] = \frac{1}{15};\ \omega^2 = Var\left[\int_0^1 V_2(r)dr\right] = \frac{1}{6300}.$$

The panel stationarity statistic thus constructed can be also shown to satisfy the same asymptotic properties as given above.

3.2 General Case with Serially Correlated Errors

In this section we consider a more general case where the stationary disturbances in the model are serially correlated with different serial correlation patterns across groups.

Assumption 3.2 *The* u_{it} *in (2.2) follow stationary AR(p) processes with heterogenous coefficients* $|\rho_i| < 1$*,*

$$u_{it} = \sum_{j=1}^{p} \rho_{ij} u_{i,t-j} + \varepsilon_{it},\ i = 1, ..., N; t = 1, ..., T, \tag{3.8}$$

where the ε_{it}*'s are iid normal variates with zero means and finite heterogenous variances,* $\sigma^2_{\varepsilon_i} > 0$.

Notice that the long run variance of u_{it} is now given by

$$\sigma^2_{u_i} = \frac{\sigma^2_{\varepsilon_i}}{\left(1 - \sum_{j=1}^{p} \rho_{ij}\right)^2}. \tag{3.9}$$

One straightforward parametric approach to estimate the long-run variance of u_{it} is given by

$$\hat{\sigma}^2_{iT,P1} = \hat{\sigma}^2_{e_iT} \times \frac{1 - \sum_{j=1}^{p} \hat{\rho}^2_{ij}}{\left(1 - \sum_{j=1}^{p} \hat{\rho}_{ij}\right)^2}, \tag{3.10}$$

where $\hat{\rho}_{ij}$ is the $\sqrt{T}$-consistent estimator of ρ_{ij} in (3.8),

$$\hat{\sigma}^2_{e_iT} = \frac{1}{T}\sum_{t=1}^{T}\hat{e}^2_{it}, \tag{3.11}$$

and $\hat{e}_{it}$ is the residual obtained from the OLS regression of y_{it} on constant. Therefore, we define the following statistic:

$$\tau_{NT,P1} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\eta_{iT,P1}-\mu}{\omega}\right), \tag{3.12}$$

where $\eta_{iT,P1} = \frac{T^{-2}\sum_{t=1}^{T}s^2_{it}}{\hat{\sigma}^2_{iT,P1}}$ is the individual statistic accommodated to deal with the presence of serial correlation.

Alternatively, we could estimate the long-run variance by

$$\hat{\sigma}^2_{iT,P2} = \frac{\hat{\sigma}^2_{\varepsilon_iT}}{\left(1-\sum_{j=1}^{p}\hat{\rho}_{ij}\right)^2}, \tag{3.13}$$

where $\hat{\sigma}^2_{\varepsilon_iT} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}^2_{it}$ and $\hat{\varepsilon}_{it}$ is the residual obtained from the OLS regression of y_{it} on constant and $y_{it-1},...,y_{i,t-p}$, and thus obtain the following statistic:

$$\tau_{NT,P2} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\eta_{iT,P2}-\mu}{\omega}\right), \tag{3.14}$$

where $\eta_{iT,P2} = \frac{T^{-2}\sum_{t=1}^{T}s^2_{it}}{\hat{\sigma}^2_{iT,P2}}$. Except for this, the algebraic representation of the test statistic is basically unchanged from the simple case.

To obtain consistent estimates of ρ_{ij}, $j = 1,...,p$, to be used in (3.9), we follow Leybourne and McCabe (1998) and estimate an over-differenced $ARIMA\,(p,1,1)$ model for the demeaned y_{it}, *i.e.*

$$\Delta\tilde{y}_{it} = \sum_{j=1}^{p}\rho_{ij}\Delta\tilde{y}_{i,t-1} + w_{it} - \theta_i w_{i,t-1}, \tag{3.15}$$

where $\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T}y_{it}$. Here we use the following generalised least squares (GLS) estimator,

$$\hat{\boldsymbol{\rho}}_i = \left(\Delta\mathbf{Z}_i^{f\prime}\Delta\mathbf{Z}_i^f\right)^{-1}\Delta\mathbf{Z}_i^{f\prime}\Delta\mathbf{y}_i^f, \tag{3.16}$$

where $\boldsymbol{\rho}_i = (\rho_{i1},\rho_{i2},...,\rho_{ip})'$ is the $p\times 1$ vector, and $\mathbf{Z}_i^f = \left(\tilde{\mathbf{y}}^f_{i,-1},\tilde{\mathbf{y}}^f_{i,-2},...,\tilde{\mathbf{y}}^f_{i,-p}\right)$ is the $T\times p$ matrix containing T filtered observations on the p lagged de-meaned series, $\tilde{\mathbf{y}}_i = (\tilde{y}_{i1},...,\tilde{y}_{iT})'$

and $\tilde{\mathbf{y}}_{i,-j} = (\tilde{y}_{i,-j+1}, ..., \tilde{y}_{iT-j})'$ with $\tilde{y}_{i,t-j} = y_{i,t-j} - \frac{1}{T}\sum_{t=1}^{T} y_{i,t-j}$. Superscript f denotes application of the filter $\left(1-\hat{\theta}_i L\right)^{-1}$ such that

$$\Delta\tilde{y}_{it}^{f} = \sum_{j=0}^{t} \hat{\theta}_i^j \Delta y_{it-j},$$

and $\hat{\theta}_i$ is the consistent estimator of the moving average parameter from the $ARIMA\,(p,1,1)$ model. We now assume:

Assumption 3.3 *For all $i = 1, ..., N$, $\boldsymbol{\rho}_i \in \Theta_\rho$ and $\sigma_{\varepsilon_i}^2 \in \Theta_\sigma$, where Θ_ρ is a compact subset of $\mathbb{R}^p$ and Θ_σ a compact subset of $\mathbb{R}$. In addition , $\boldsymbol{\rho}_i$ and $\sigma_{\varepsilon_i}^2$ $i = 1, 2...N$ are deterministic parameter sequences.*

Again following Leybourne and McCabe (1998), it can be shown under Assumptions 3.2 and 3.3 that the GLS estimator $\hat{\boldsymbol{\rho}}_i$ is $\sqrt{T}$-consistent for both cases with $\theta_i = 1$ (null) and with $\theta_i < 1$ (alternative).

Using the sequential asymptotic approach, it is now straightforward to show that both mean group statistics, $\tau_{NT,P1}$ and $\tau_{NT,P2}$, weakly converge to standard normal variates. First, for each i, as $T \to \infty$, the individual statistics $\eta_{iT,P1}$ and $\eta_{iT,P2}$ weakly converge to η_i as defined by (3.3), since both $\hat{\sigma}_{iT,P1}^2$ and $\hat{\sigma}_{iT,P2}^2$ are consistent estimates of the long-run variance of u_{it} given by (3.9). That is, as $T \to \infty$,

$$\tau_{NT,P1} \overset{T}{\Rightarrow} \tau_N,\ \tau_{NT,P2} \overset{T}{\Rightarrow} \tau_N, \tag{3.17}$$

where $\tau_N = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\eta_i-\mu}{\omega}\right)$. Again noting that η_i's are *iid* with mean μ and variance ω^2 and using the Lindberg-Levy central limit theorem, we also have as $N \to \infty$, $\tau_N \overset{N}{\Rightarrow} N(0,1)$.

The prediction of sequential asymptotic theory is again that the mean group statistics $\tau_{NT,P1}$ and $\tau_{NT,P2}$ will be asymptotically standard normal. Next, we will derive a further condition on N and T for the standard normal limit to hold under joint asymptotic analysis.

Theorem 3.2 *Suppose that Assumptions 3.2 and 3.3 hold. Then, under the null hypothesis (2.3), as $T \to \infty$ and $N \to \infty$ with $N/T \to 0$, the panel-based stationarity test statistic defined by (3.12) or (3.14) weakly converges to a standard normal variate. Under the alternative hypothesis (2.4), as $T, N \to \infty$ and $N_1/N = \delta > 0$, it diverges to infinity.*

Proof. See the Appendix. ■

Theorem 3.2 also shows that it is necessary for the validity of the test under the null that N/T tends to zero. But, notice that this finding is consistent with Phillips and Moon (1999) who also find that sequential asymptotic results can only be extended to joint results with an additional condition on the relative size of T and N.

4 Monte Carlo Simulation Results

In this section we use Monte Carlo experiments to examine finite sample properties of our proposed panel-based stationarity test.

We consider the two sets of Monte Carlo experiments. The first set focuses on the benchmark model,

$$y_{it} = \alpha_i + e_{it},\ i = 1, ..., N;\ t = 1, ...T, \tag{4.1}$$

$$e_{it} = \gamma_{it} + u_{it},\ \gamma_{it} = \gamma_{it-1} + v_{it},\ \gamma_{i0} = 0. \tag{4.2}$$

The second set of experiments allows for the presence of positive (heterogeneous) AR(1) serial correlations in u_{it},

$$u_{it} = \rho_i u_{i,t-1} + \varepsilon_{it},\ t = 1, ..., T;\ i = 1, ..., N, \tag{4.3}$$

where $\rho_i \sim U[0.2, 0.4]$, U stands for a uniform distribution and ρ_i's are generated independently of ε_{it}.

In all of the experiments ε_{it} (or u_{it} when $\rho_i = 0$) and v_{it} are independently generated as *iid* normal variates with zero means and heterogeneous variances, $\sigma^2_{u_i}$ and $\sigma^2_{v_i}$. The parameters α_i, $\sigma^2_{u_i}$ and $\sigma^2_{v_i}$ are generated by

$$\alpha_i \sim N(0, 1),\ \sigma^2_{u_i} \sim U[0.5, 1.5],\ \sigma^2_{v_i} \sim \sigma^2_v \times U[0.5, 1.5],\ i = 1, 2, ..., N. \tag{4.4}$$

All of the parameter values such as α_i, $\sigma^2_{u_i}$, $\sigma^2_{v_i}$, ρ_i or ψ_i are generated independently of ε_{it} and v_{it} once, and then fixed throughout replications. Throughout all of Monte Carlo experiments we set $\sigma^2_v = 0$ under the null, but $\sigma^2_v = 0.01$ (Experiment 1) and $\sigma^2_v = 0.1$ (Experiment 2) under the alternative hypothesis. We will evaluate empirical size and power of the alternative tests at 5% nominal level.

The first set of experiments based on 5,000 replications were carried out for $T = 15$, 20, 30, 50, 100, and $N = 1, 10, 25, 50, 100$. Here we only consider the τ_{NT} statistic defined in (3.5), and the stochastic simulation results are summarized in Table 1(a).

Table 1(*a*) about *here*

As a benchmark, we also give the results for $N = 1$. The simulation results here clearly show that the τ_{NT} test performs well when T is large relative to N. The power of the test rises monotonically with N and T, though it depends critically more on T. Hence, it is possible to substantially augment the power of the stationarity tests applied to single time series. Turning to the size performance, we find that the τ_{NT} test tends to over-reject in some situations, especially as N increases relative to T, though the size of the τ_{NT} test gets closer to the nominal 5% as T increases for a fixed N. Overall this finding is quite consistent with the joint asymptotic result in Section 3 that the normal approximation of the τ_{NT} test is valid only under $N/T \to 0$.

In general, the above result is unsatisfactory in the sense that there is no clear guidance on the ratio of N to T in practice that avoids substantial upward size distortions. To remedy these problems we now suggest the following finite-sample adjustment, which is designed for the pseudo panels where $T \geq N$.[1] Consider the non-parametric group-based estimator of ω^2, denoted by $\hat{\omega}_N^2$ and obtained simply by

$$\hat{\omega}_N^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(\eta_{iT} - \overline{\eta_T}\right)^2 ,$$

where η_{iT} is the individual stationarity statistic defined by (3.1), and $\overline{\eta_T} = \frac{1}{N}\sum_{i=1}^{N} \eta_{iT}$ is the empirical group mean. We now propose the following modified statistic:

$$\tau_{NT}(\delta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\eta_{iT} - \mu\right) \times \frac{1}{\omega^{\delta}\hat{\omega}_N^{1-\delta}},\ 0 \leq \delta \leq 1. \tag{4.5}$$

$\tau_{NT}(\delta)$ can also be rewritten as

$$\tau_{NT}(\delta) = \left(\frac{\omega}{\hat{\omega}_N}\right)^{1-\delta} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\left(\eta_{iT} - \mu\right)}{\omega} = \left(\frac{\omega}{\hat{\omega}_N}\right)^{1-\delta} \tau_{NT}.$$

The asymptotic equivalence of the modified $\tau_{NT}(\delta)$ test to the τ_{NT} test under the null is clearly ensured since $\frac{\omega}{\hat{\omega}_N} \to 1$ under the null as T and N both tend to grow without bound. Under the alternative, it is easily seen that the test diverges at rate $\sqrt{N}T^{2\delta}$.

The first set of experiments based on 5,000 replications were repeated using the $\tau_{NT}(\delta)$ test for different values of δ. Table 1(b) now presents these simulation results for $\delta = 1, 0.75, 0.5, 0.25, 0$.

Table 1(*b*) about *here*

For purposes of comparison we also give the result for $\delta = 1$, which is the same as the unmodified τ_{NT} test. Table 1(b) clearly shows that the rejection frequency of the test always gets smaller as δ decreases. For example, the test is over-sized when $\delta = 1$, while it is under-sized in all most all cases when $\delta = 0$. In general, the direct power comparison of the $\tau_{NT}(\delta)$ for different values of δ is problematic, and thus the significantly larger power of the $\tau_{NT,P1}(1)$ test over other tests should be discounted. For this reason, we recommend to use the $\tau_{NT,P1}(0.5)$ test in practice, since its size is close to nominal level in most cases, in particular when $T \geq N$. This is clearly a compromise, since choosing higher value of δ would give over-rejection (size distortion), while choosing lower value of δ would make the test less powerful. Comparing the power of the test based on the single time series (see Table 1(a)), we still find that there is a substantial gain in the power of the panel-based stationarity tests.

Next, we consider the Experiment 2 where the underlying DGP contain serially correlated errors. In this case we should use the test developed in section 3.2, namely $\tau_{NT,P1}$ and $\tau_{NT,P2}$ statistics defined by (3.12) and (3.14). We also consider the finite sample modification used

[1] Notice that the finite sample correction suggested here is not clearly plausible in the single time series case. Though we focus on the case of $T \geq N$, this correction seems to work in the case where $T \geq 30$ and N is slightly larger than T as will be shown.

previously, i.e. see (4.5), and denote them as $\tau_{NT,P1}(\delta)$ and $\tau_{NT,P2}(\delta)$, accordingly. As will be shown below this modification plays a more important role since the unmodified test will suffer from more size distortions in the presence of serial correlation.

In addition we consider alternative panel stationarity tests based on the non-parametric estimation of the underlying long-run variance as suggested by KPSS in the single time series; that is,

$$\tau_{NT,NP(\ell)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{\eta_{iT,NP(\ell)} - \mu}{\omega} \right), \tag{4.6}$$

where

$$\eta_{iT,NP(\ell)} = \frac{S_{iT}}{s_i^2(\ell)};\ s_i^2(\ell) = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{it}^2 + \frac{2}{T}\sum_{j=1}^{\ell} w(j,\ell) \sum_{t=1}^{T} \hat{e}_{it}\hat{e}_{it-j},$$

and $w(j,\ell)$ is an optional weighting function that corresponds to the choice of a spectral window. Following KPSS we here use the Bartlett window, $w(j,\ell) = 1 - \frac{j}{\ell+1}$. As was shown by KPSS the choice of the lag truncation, ℓ, is crucial for the test to have the reasonable finite sample performance, though there is no simple choice in practice. Here we use the two lag truncation values $\ell_4 = \text{integer}\left[4 \times \left(\frac{T}{100}\right)^{\frac{1}{4}}\right]$ and $\ell_{12} = \text{integer}\left[12 \times \left(\frac{T}{100}\right)^{\frac{1}{4}}\right]$, following KPSS. Though we have not provided any theoretical justification for the validity of such tests, a sketch of their properties may be gained via Monte Carlo experiments.

Tables 2(a)-2(c) summarize the simulation results of alternative tests for the second set of experiments .

$$Tables\ 2(a) - 2(c)\ \text{about}\ here$$

First, the results in Table 2(a) clearly show the importance of appropriately choosing the value of δ for controlling the size performance of the tests in general. For example, when $\delta = 1$ and thus no finite sample modification is made, then the $\tau_{NT,P1}$ test suffers from the over-rejection. Its size tends to nominal level as T increases for a fixed N, but quite slowly. The problem is particularly more serious as N increases for a fixed T. Again this behavior is consistent with the joint asymptotic theory. On the other hand, the choice of $\delta = 0$ would render the test being under-sized in all cases. Notice also that the size of the $\tau_{NT,P1}(0)$ test seems to slowly tend to nominal size as N increases for a fixed T. As expected, the higher the value of δ is, the more the rejection frequency of the test and *vice versa*. Turning to the power performance, the power of the tests rises monotonically with N and T, but the rise in T improves the power more significantly. One important finding with the serially correlated data is that we need sufficiently large time periods to augment the power of the panel-based stationarity tests. For example, when $T = 15$, the power of the $\tau_{NT,P1}(0.5)$ test increases only from 0.042 for $N = 10$ to 0.112 for $N = 100$, while when $T = 20$, the power of the $\tau_{NT,P1}(0.5)$ test increases from 0.147 for $N = 10$ to 0.747 for $N = 75$. Against the alternative used here we find that the power of the $\tau_{NT,P1}(0.5)$ test is quite substantial when $T \geq 30$, which is clearly a great improvement over the single time series case. This finding leads us to recommend the use of the $\tau_{NT,P1}(0.5)$ test in practice, the size of which is more or less close to the nominal 5% level in most cases when $T \geq N$. Alternatively, when $N \geq T$, but N is not extremely larger than T, then the $\tau_{NT,P1}(0.25)$ test may be considered.

Second, similar phenomena have been found if we consider the results for the $\tau_{NT,P2}$ test presented in Table 2(b), though the size-distortion of the $\tau_{NT,P2}(1)$ test is much more serious than the $\tau_{NT,P1}(1)$ test. Based on the overall simulation results in this case we may recommend use of the $\tau_{NT,P2}(0.25)$ test, though it has a slightly worse size-distortion than the $\tau_{NT,P1}(0.5)$ test, as N rises for a given T. Turning to the power comparison between the $\tau_{NT,P1}(0.5)$ and $\tau_{NT,P2}(0.25)$ tests, we find that the power of both tests are similar.

Finally, the simulation results for the finite sample performance of the tests based on the non-parametric estimation of the underlying long-run variance are presented in Table 2(c). Here we find quite strongly that the non-parametric correction in the panel context does not seem to work properly. In particular, there are massive size distortions for almost all cases considered. Even the finite sample modification, which has been effective to controlling size of the test based on the parametric estimation of the long-run variance, does not reduce the size distortion to any significant degree. Looking at the results for $\tau_{NT,NP(\ell_{12})}(1)$ more closely, the size continues to be closer to the nominal level as T increases for a fixed N, e.g. when $N = 10$, from .899 for $T = 10$ to .113 for $T = 100$. But, this rate of convergence seems to be much slower than in the single time series case. Overall this simulation result may indicate that the use of non-parametric based statistic is much less useful in the panels with the typical sample sizes encountered in practice.

In sum, the Monte Carlo results seem to support the joint asymptotic theory. Moreover, the finite sample modification suggested in this section seems clearly effective in rendering the size of the tests based on the parametric estimation of the long–run variance close to the nominal level while retaining significantly higher power than obtained in the single time series case. Considering that it is more important to have the correct nominal size in the setting of the stationarity test, we recommend to use the $\tau_{NT,P1}(0.5)$ test in empirical application though there may be some situations where this test tend to over-reject especially when $T \leq N$.[2]

5 Concluding Remarks

In this paper we have developed a computationally simple procedure for testing the null of stationarity hypothesis against the alternative of unit roots in heterogeneous pseudo panels where N and T are jointly large. Using the joint asymptotic approach we have shown that the suggested test statistic has a standard normal distribution as both N and T grow without bounds but with N/T tending to zero under the null, even in the case where panels have both parameters heterogeneity across cross-section units and serial correlation across time periods.

The small sample properties of the proposed tests are investigated via Monte Carlo methods. It is found that when there are no serial correlations in the underlying errors,

[2]We have also examined a third set of experiments which allows for a linear trend in estimation of the ADF regressions using the same data generating process employed in the second set of experiments. The simulation results are qualitatively similar to the above except the power increases more slowly with increased N than before.

the suggested tests perform very well even as T is small as 15. In this case it is possible to substantially augment the power of the stationarity tests applied to single time series. The situation is more complicated when the disturbances in the panel are serially correlated; in fact, the suggested test tends to over-reject especially when T is relatively small. This over-size problem is shown to be greatly reduced by a finite-sample adjustment based on the empirically obtained non-parametric group-based estimator of the individual stationarity statistics. This modification seems to be very effective in the pseudo panels where $T \geq N$, while at the same time maintaining the substantial power boost.

Extension of the test to the case where disturbances are correlated across panel units presents no great difficulties provided these correlations may be satisfactorily accommodated by a simple common time specific additive factor in each panel equation. In this case we could follow Im *et al.* (2002) and work with the cross-section de-meaned data, $y_{it} - \overline{y_t}^g$, where $\overline{y_t}^g = \sum_{i=1}^{N} y_{it}/N$. The proofs can be readily but tediously extend to such a case. Considering that there are currently a few studies investigating the impact of a general structure of cross section dependence on the performance of unit root tests in dynamic heterogeneous panels [*e.g.* Chang (2000) and Phillips and Sul (2002)], however, the extension of the current mean group tests for stationarity along this line of research will be of more interest, which we leave for future research.

Table 1(a). Finite Sample Performance of the τ_{NT} Test for Experiment 1										
$T\backslash N$	1		10		25		50		100	
	size	power	size	power	size	power	size	power	size	power
15	.056	.083	.082	.156	.083	.254	.111	.422	.165	.616
20	.055	.117	.072	.413	.072	.717	.086	.920	.109	.989
30	.049	.146	.073	.484	.071	.843	.083	.941	.094	.999
50	.050	.287	.067	.927	.067	.996	.080	1.00	.073	1.00
100	.049	.587	.061	1.00	.066	1.00	.063	1.00	.066	1.00

Table 1(b). Finite Sample Performance of the $\tau_{NT}(\delta)$ Test for Experiment 1											
δ		1		0.75		0.5		0.25		0	
T	N	size	power	size	power	size	power	size	power	size	power
15	10	.082	.156	.071	.137	.061	.121	.051	.104	.043	.088
20	10	.068	.270	.057	.241	.045	.207	.036	.171	.032	.140
30	10	.073	.484	.062	.446	.046	.402	.033	.331	.026	.247
50	10	.067	.927	.053	.914	.042	.892	.029	.846	.021	.713
100	10	.061	1.00	.049	1.00	.036	1.00	.025	1.00	.017	.986
15	25	.083	.254	.076	.238	.067	.222	.061	.207	.056	.189
20	25	.085	.438	.077	.412	.067	.393	.060	.364	.052	.325
30	25	.071	.843	.062	.823	.048	.797	.042	.758	.035	.686
50	25	.067	.996	.057	.995	.047	.994	.036	.990	.028	.975
100	25	.066	1.00	.056	1.00	.044	1.00	.031	1.00	.020	1.00
15	50	.111	.422	.107	.412	.104	.402	.099	.390	.095	.381
20	50	.095	.674	.090	.662	.085	.645	.079	.622	.073	.594
30	50	.083	.941	.077	.936	.069	.927	.063	.914	.056	.894
50	50	.080	1.00	.068	1.00	.059	1.00	.049	1.00	.041	1.00
100	50	.063	1.00	.053	1.00	.045	1.00	.038	1.00	.030	1.00
15	100	.165	.616	.167	.616	.168	.617	.170	.616	.172	.616
20	100	.124	.902	.120	.900	.117	.897	.113	.893	.108	.887
30	100	.094	.999	.087	.999	.080	.999	.076	.999	.070	.999
50	100	.073	1.00	.067	1.00	.061	1.00	.054	1.00	.047	1.00
100	100	.066	1.00	.060	1.00	.053	1.00	.046	1.00	.040	1.00

Table 2(a). Finite Sample Performance of the $\tau_{NT,P1}(\delta)$ Test for Experiment 2

δ		1		0.75		0.5		0.25		0	
T	N	size	power	size	power	size	power	size	power	size	power
15	10	.066	.082	.041	.061	.030	.042	.014	.025	.006	.015
20	10	.104	.230	.070	.190	.049	.147	.025	.094	.009	.048
30	10	.118	.614	.093	.567	.064	.502	.033	.380	.011	.201
50	10	.112	.928	.088	.918	.058	.895	.028	.833	.011	.634
100	10	.084	1.00	.064	1.00	.046	1.00	.025	1.00	.015	.990
15	25	.056	.098	.042	.080	.028	.059	.019	.045	.010	.030
20	25	.105	.385	.085	.349	.064	.296	.038	.228	.017	.151
30	25	.153	.872	.128	.856	.087	.823	.051	.753	.022	.627
50	25	.132	1.00	.106	1.00	.076	1.00	.044	.999	.021	.993
100	25	.084	1.00	.068	1.00	.051	1.00	.035	1.00	.018	1.00
15	50	.055	.141	.044	.122	.033	.101	.020	.085	.013	.059
20	50	.118	.605	.096	.576	.073	.520	.050	.450	.023	.371
30	50	.174	.982	.146	.980	.110	.976	.074	.961	.038	.927
50	50	.136	1.00	.106	1.00	.083	1.00	.053	1.00	.032	1.00
100	50	.077	1.00	.064	1.00	.049	1.00	.035	1.00	.024	1.00
15	100	.034	.150	.028	.132	.021	.112	.015	.096	.009	.077
20	100	.130	.805	.106	.783	.090	.747	.068	.710	.043	.660
30	100	.208	1.00	.173	1.00	.132	.999	.100	.999	.068	.999
50	100	.148	1.00	.124	1.00	.098	1.00	.075	1.00	.032	1.00
100	100	.096	1.00	.081	1.00	.065	1.00	.052	1.00	.038	1.00

Table 2(b). Finite Sample Performance of the $\tau_{NT,P2}(\delta)$ Test for Experiment 2

δ		1		0.75		0.5		0.25		0	
T	N	size	power	size	power	size	power	size	power	size	power
15	10	.088	.152	.065	.113	.040	.082	.017	.045	.006	.018
20	10	.154	.356	.113	.306	.068	.234	.036	.138	.007	.048
30	10	.179	.721	.138	.697	.088	.637	.043	.479	.010	.166
50	10	.155	.952	.120	.945	.082	.932	.035	.882	.011	.523
100	10	.104	1.00	.087	1.00	.061	1.00	.028	1.00	.014	.888
15	25	.091	.198	.068	.170	.048	.128	.025	.080	.013	.043
20	25	.178	.570	.140	.521	.100	.447	.051	.340	.015	.175
30	25	.228	.932	.196	.922	.146	.897	.072	.837	.022	.610
50	25	.202	1.00	.163	1.00	.114	1.00	.059	1.00	.021	.974
100	25	.117	1.00	.090	1.00	.068	1.00	.043	1.00	.019	1.00
15	50	.087	.270	.071	.227	.049	.189	.032	.140	.015	.095
20	50	.192	.796	.150	.760	.112	.701	.072	.606	.033	.468
30	50	.287	.995	.239	.993	.182	.989	.111	.983	.046	.947
50	50	.229	1.00	.187	1.00	.139	1.00	.081	1.00	.038	1.00
100	50	.118	1.00	.095	1.00	.069	1.00	.045	1.00	.025	1.00
15	100	.082	.344	.062	.303	.045	.262	.031	.210	.016	.152
20	100	.239	.948	.193	.936	.153	.914	.109	.880	.062	.798
30	100	.327	1.00	.281	1.00	.232	1.00	.166	1.00	.089	1.00
50	100	.275	1.00	.226	1.00	.179	1.00	.125	1.00	.069	1.00
100	100	.150	1.00	.131	1.00	.106	1.00	.071	1.00	.045	1.00

Table 2(c). Finite Sample Performance of Non-parametric-based Tests for Experiment 2													
		$\tau_{NT,NP(\ell_4)}(\delta)$						$\tau_{NT,NP(\ell_{12})}(\delta)$					
δ		1		0.5		0		1		0.5		0	
T	N	size	power	size	power	size	power	size	power	size	power	size	power
15	10	.365	.787	.411	.814	.467	.825	.899	.963	1.00	.999	1.00	1.00
20	10	.424	.877	.435	.875	.454	.850	.636	.876	.907	.984	.989	1.00
30	10	.293	.957	.301	.947	.311	.921	.261	.821	.496	.917	.718	.967
50	10	.350	.998	.314	.995	.267	.985	.131	.935	.200	.942	.292	.951
100	10	.296	1.00	.258	1.00	.195	1.00	.113	1.00	.116	1.00	.130	.998
15	25	.788	.993	.890	.996	.937	.997	1.00	1.00	1.00	1.00	1.00	1.00
20	25	.783	.998	.811	.999	.839	.999	.922	1.00	1.00	1.00	1.00	1.00
30	25	.611	1.00	.646	1.00	.675	1.00	.738	.999	.924	1.00	.982	1.00
50	25	.640	1.00	.635	1.00	.624	1.00	.324	1.00	.469	1.00	.606	1.00
100	25	.486	1.00	.455	1.00	.417	1.00	.172	1.00	.193	1.00	.233	1.00
15	50	.960	1.00	.980	1.00	.990	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	50	.968	1.00	.979	1.00	.985	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30	50	.898	1.00	.923	1.00	.945	1.00	.982	1.00	.999	1.00	1.00	1.00
50	50	.884	1.00	.890	1.00	.897	1.00	.610	1.00	.765	1.00	.873	1.00
100	50	.779	1.00	.775	1.00	.760	1.00	.291	1.00	.355	1.00	.428	1.00
15	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
20	100	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30	100	.994	1.00	.998	1.00	.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50	100	.997	1.00	.997	1.00	.999	1.00	.921	1.00	.978	1.00	.995	1.00
100	100	.956	1.00	.959	1.00	.960	1.00	.555	1.00	.632	1.00	.709	1.00

A Appendix A

In what follows the phrase "k-th moment of z_{it} is bounded" will be used to imply that for any given i and T, $E(z_{it}^k)$ has some fixed upper bound and for any given i this upper bound is $O(1)$ (or $o(1)$) in T.

A.1 Proof of Theorem 3.1

For this we analyse the discrepancy between τ_{NT} and $\tau_{N\infty}$; that is, analyse D_{NT} where

$$D_{NT} = \tau_{N\infty} - \tau_{NT} = \frac{1}{\omega}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} d_{iT}, \tag{A.1}$$

where

$$d_{iT} = \eta_{i\infty} - \eta_{iT} = \frac{S_{i\infty}}{\sigma_i^2} - \frac{S_{iT}}{\hat{\sigma}_{iT}^2} = \frac{(S_{i\infty} - S_{iT})}{\sigma_i^2} + \frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2}, \tag{A.2a}$$

$$S_{iT} = T^{-2}\sum_{t=1}^{T} s_{it}^2;\ \eta_{iT} = \frac{S_{iT}}{\hat{\sigma}_{iT}^2};\ \eta_{i\infty} = \frac{S_{i\infty}}{\sigma_i^2}. \tag{A.3}$$

Using (A.1) and (A.2a) we may write

$$D_{NT} = \frac{1}{\omega}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{(S_{i\infty} - S_{iT})}{\sigma_i^2} + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right\} = \frac{1}{\omega}\left(A_{NT} + B_{NT}\right). \tag{A.4}$$

Then using the fact that $\tau_{N\infty}$ converges in distribution to a standard normal variate, it is sufficient for the proof of the theorem to show that under the null as $T \to \infty$ and $N \to \infty$ with $N/T \to 0$,

$$\text{plim}_{N,T\to\infty,N/T\to 0} D_{NT} = 0. \tag{A.5}$$

We examine the probability limits of A_{NT} and B_{NT} in (A.4), separately. Notice that under $\sigma_{v_i}^2 = 0$,

$$y_{it} = \alpha_i + u_{it}, \tag{A.6}$$

and thus $\sigma_i^2 = \sigma_{u_i}^2$.

We now show that the mean and variance of $\frac{S_{i\infty} - S_{iT}}{\sigma_i^2}$ are bounded functions of T alone and are $O(T^{-1})$ and $o(1)$ in T, respectively. We start by showing that all moments of S_{iT} exist. By definition we have

$$T^2 S_{iT} = (u_{i1}, u_{i2}, ..., u_{iT})\,\mathbf{S}\,(u_{i1}, u_{i2}, ..., u_{iT})', \tag{A.7}$$

$$\mathbf{S} = \mathbf{MPP'M}, \tag{A.8}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{i}_T\left(\mathbf{i}_T'\mathbf{i}_T\right)^{-1}\mathbf{i}_T'$, $\mathbf{i}_T$ denote a $T\times 1$ vector of units and $\mathbf{P}$ is a $T\times T$ upper triangular matrix of units. Writing $\mathbf{S}$ in its canonical form as $\mathbf{E}\Lambda\mathbf{E}'$ where $\Lambda = diag\left(\lambda_j\right)$ are the eigenvalues of $\mathbf{S}$, and $\mathbf{E}$ contains the corresponding orthonormal eigenvectors, gives

$$T^2 S_{iT} = \sum_{t=1}^{T}\lambda_t u_{it}^{*\,2}, \tag{A.9}$$

where $(u_{i1}^*, u_{i2}^*, ..., u_{iT}^*) = (u_{i1}, u_{i2}, ..., u_{iT})\,\mathbf{E}$ is a $T\times 1$ vector of independent $N\left(0, \sigma_i^2\right)$ variates. Using Minkowski's inequality we have

$$\begin{aligned}\left[E\left(\sum_{t=1}^{T}\lambda_t u_{it}^{*\,2}\right)^r\right]^{\frac{1}{r}} &\leq \sum_{t=1}^{T}\left\{E\left[\left(\lambda_t u_{it}^{*\,2}\right)^r\right]\right\}^{\frac{1}{r}} = \sum_{t=1}^{T}\lambda_t\left[E\left(u_{it}^{*\,2r}\right)\right]^{\frac{1}{r}} \\ &= 2\sigma_i^2(2r!)^{\frac{1}{2r}}\sum_{t=1}^{T}\lambda_t = 2\sigma_i^2(2r!)^{\frac{1}{2r}} \times \frac{T^3 - 9T^2 + 5T - 3}{6T},\end{aligned} \tag{A.10}$$

where the last equality follows from the fact that $\sigma_i^2 \sum_{t=1}^{T} \lambda_t = E(T^2 S_{iT})$. Dividing (A.10) by $\sigma_i^2 T^2$ shows that the rth moment of $\frac{S_{iT}}{\sigma_i^2}$ has an upper bound that is $o(1)$ in T. This allows us to invoke the Corollary of Theorem 25.12 of Billingsley(1979) and interchange the limit and integration(expectation) operators [3] and write

$$\lim_{T\to\infty} E\left[\left(\frac{S_{iT}}{\sigma_i^2}\right)^r\right] = E\left[\left(\frac{S_{i\infty}}{\sigma_i^2}\right)^r\right]. \tag{A.11}$$

By Hölder's inequality we also obtain

$$\lim_{T\to\infty} E\left(\frac{S_{iT}}{\sigma_i^2}\frac{S_{i\infty}}{\sigma_i^2}\right) \leq \lim_{T\to\infty} \sqrt{E\left[\left(\frac{S_{iT}}{\sigma_i^2}\right)^2\right]}\sqrt{E\left[\left(\frac{S_{i\infty}}{\sigma_i^2}\right)^2\right]} = E\left[\left(\frac{S_{i\infty}}{\sigma_i^2}\right)^2\right] \leq 2(4!)^{\frac{1}{4}}. \tag{A.12}$$

Again boundedness of moments allows the interchange of limits and expectations to give

$$\lim_{T\to\infty} E\left(\frac{S_{iT}}{\sigma_i^2}\frac{S_{i\infty}}{\sigma_i^2}\right) = E\left[\left(\frac{S_{i\infty}}{\sigma_i^2}\right)^2\right]. \tag{A.13}$$

Using (A.10), (A.12) and the fact that second moment exceeds variance we have

$$\begin{aligned}\lim_{T\to\infty} Var\left(\frac{S_{iT}-S_{i\infty}}{\sigma_i^2}\right) &\leq \lim_{T\to\infty} E\left[\left(\frac{S_{iT}-S_{i\infty}}{\sigma_i^2}\right)^2\right] \\ &= \lim_{T\to\infty} E\left[\left(\frac{S_{iT}}{\sigma_i^2}\right)^2\right] + \lim_{T\to\infty} E\left[\left(\frac{S_{i\infty}}{\sigma_i^2}\right)^2\right] - 2\lim_{T\to\infty} E\left(\frac{S_{iT}S_{i\infty}}{\sigma_i^2\sigma_i^2}\right) = 0.\end{aligned}$$

Expanding $E\left(S_{iT}\right)$ gives

$$E\left(\frac{S_{iT}}{\sigma_i^2}\right) = \frac{1}{6} - \frac{9T^2-5T+3}{6T^3} = \lim_{T\to\infty} E\left(\frac{S_{iT}}{\sigma_i^2}\right) - \frac{9T^2-5T+3}{6T^3} = E\left(\frac{S_{i\infty}}{\sigma_i^2}\right) + O(T^{-1}). \tag{A.14}$$

(A.14) and (A.14) establish that

$$\lim_{T\to\infty} Var\left(\frac{S_{iT}-S_{i\infty}}{\sigma_i^2}\right) = 0;\ E\left(\frac{S_{i\infty}-S_{iT}}{\sigma_i^2}\right) = O(T^{-1}).$$

Denoting the first and second moment of $\frac{S_{i\infty}-S_{iT}}{\sigma_i^2}$ for some fixed T (and any i) as $M(T)$ and $V(T)$ respectively, it follows that

$$E\left(A_{NT}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{E(S_{i\infty}-S_{iT})}{\sigma_i^2} = E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M(T)\right] = N^{\frac{1}{2}}M(T) = O\left(\frac{\sqrt{N}}{T}\right), \tag{A.15}$$

and thus

$$\lim_{N,T\to\infty,\frac{N}{T}\to 0} E(A_{NT}) = 0. \tag{A.16}$$

Next, using the independence of panel units we may write

$$\begin{aligned}\lim_{N,T\to\infty,\frac{N}{T}\to 0} Var\left(A_{NT}\right) &= \lim_{N,T\to\infty,\frac{N}{T}\to 0} Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{S_{i\infty}-S_{iT}}{\sigma_i^2}\right) = \lim_{N,T\to\infty,\frac{N}{T}\to 0}\frac{1}{N}\sum_{i=1}^{N} V(T) \\ &= \lim_{N,T\to\infty,\frac{N}{T}\to 0} V(T) = \lim_{T\to\infty} o(1) = 0.\end{aligned} \tag{A.17}$$

[3] This follows by application of the dominated convergence theorem. See, for example, Billingsley(1979, p. 180).

Using Chebyshev's inequality, (A.16) and (A.17) imply that

$$\text{plim}_{N,T\to\infty,N/T\to 0}A_{NT} = \text{plim}_{N,T\to\infty,N/T\to 0}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{S_{i\infty}-S_{iT}}{\sigma_i^2}\right) = 0. \tag{A.18}$$

We now show that the first and second moments of B_{NT} in (A.4) are bounded and are $O\left(T^{-\frac{1}{2}}\right)$ and $o(1)$ in T, respectively. Regarding the second moment of B_{NT} we have

$$\begin{aligned} E\left[\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right]^2 &= \frac{1}{T}E\left\{\left(\frac{S_{iT}}{\sigma_i^2}\times\frac{1}{\hat{\sigma}_{iT}^2}\right)^2\left[\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)\right]^2\right\} \\ &\le \frac{1}{T}\sqrt{E\left(\frac{S_{iT}}{\sigma_i^2}\times\frac{1}{\hat{\sigma}_{iT}^2}\right)^4 E\left[\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)\right]^4}, \end{aligned}$$

where the second line follows from Hölder's inequality. Applying Hölder's inequality to $E\left(\frac{S_{iT}}{\sigma_i^2}\times\frac{1}{\hat{\sigma}_{iT}^2}\right)^4$ gives

$$E\left(\frac{S_{iT}}{\sigma_i^2}\times\frac{1}{\hat{\sigma}_{iT}^2}\right)^4 \le \sqrt{E\left(\frac{S_{iT}}{\sigma_i^2}\right)^8 E\left(\frac{1}{\hat{\sigma}_{iT}^2}\right)^8}. \tag{A.19}$$

Notice that we have already shown above that $E\left(\frac{S_{iT}}{\sigma_i^2}\right)^8$ is bounded (set $r=8$ in (A.10)). The second term under the square root in (A.19) is T times a univariate Inverted Wishart (IW) variable with $T-1$ degrees of freedom. By direct integration using the formula for the IW *pdf* (e.g. Press (1972, p. 111)), it is easy to show that the eighth moment of this term is finite for any i and $T > T_L$ and is $O(1)$ in T for any i.

Next, the term $\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)$ is a standard textbook quantity from a fixed regressors regression with normally distributed errors. It is well known that this quantity (which is proportional to a mean corrected χ_T^2) has a zero mean and bounded 4th moment.

Notice that all of the upper bounds given above and referred to in the previous paragraph are functions of σ_i^2 and T *only*. Because these functions depend on i only through σ_i^2 we may write

$$E\left[\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right]^2 = \frac{1}{T}E\left[\frac{S_{iT}}{\sigma_i^2}\frac{\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right]^2 \le \frac{1}{T}Q_T\left(\sigma_i^2\right). \tag{A.20}$$

Notice here that $Q_T\left(\sigma_i^2\right)$ is a bounded function for all $\sigma_i^2 \in \Theta_\sigma$ where Θ_σ is the compact set, and therefore is $O(1)$ in T. Using these properties of $Q_T(\cdot)$ we may define for $T \ge T_L$,

$$Q_T^* = \max_{\sigma_i^2\in\Theta_\sigma} Q_T\left(\sigma_i^2\right) < \infty. \tag{A.21}$$

Combining (A.20) and (A.21) and using the property of independence across panels we have

$$\begin{aligned} E\left(B_{NT}^2\right) &= \frac{1}{NT}E\left\{\sum_{i=1}^{N}\left[\frac{S_{iT}}{\sigma_i^2}\frac{\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right]\right\}^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}E\left[\frac{S_{iT}}{\sigma_i^2}\frac{\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right]^2 + \frac{2}{NT}\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}E\left[\frac{S_{jT}}{\sigma_j^2}\frac{\sqrt{T}\left(\hat{\sigma}_{jT}^2-\sigma_j^2\right)}{\hat{\sigma}_{jT}^2}\right]E\left[\frac{S_{kT}}{\sigma_k^2}\frac{\sqrt{T}\left(\hat{\sigma}_{kT}^2-\sigma_k^2\right)}{\hat{\sigma}_{kT}^2}\right] \\ &\le \frac{1}{NT}\sum_{i=1}^{N}Q_T^* + \frac{2}{NT}\sum_{j=1}^{N-1}\sum_{k=j+1}^{N}\sqrt{Q_T^*}\sqrt{Q_T^*} = \frac{1}{T}Q_T^* + \frac{(N-1)}{T}Q_T^* = \frac{N}{T}Q_T^*, \end{aligned} \tag{A.22}$$

which clearly shows that

$$\lim_{N,T\to\infty,\frac{N}{T}\to 0} E(B_{NT}^2) = 0. \tag{A.23}$$

Using Jensen's inequality it follows that

$$\lim_{N,T\to\infty,\frac{N}{T}\to 0} E(B_{NT}) = 0. \tag{A.24}$$

Chebyshev's inequality now implies that

$$\text{plim}_{N,T\to\infty,\frac{N}{T}\to 0} B_{NT} = 0. \tag{A.25}$$

From the definition of D_{NT} given in (A.4), it is clear that (A.18) and (A.25) establishes (A.5).

Next, under the alternative hypothesis (2.4) we have

$$\begin{aligned} \tau_{NT} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\eta_{iT}-\mu}{\omega} \\ &= \sqrt{\frac{N_1}{N}}\frac{1}{\omega\sqrt{N_1}}\sum_{i=1}^{N_1}\left(\frac{S_{iT,1}}{\widehat{\sigma}^2_{iT,1}}-\mu\right)+\sqrt{\frac{N_2}{N}}\frac{1}{\omega\sqrt{N_2}}\sum_{i=1}^{N_2}\left(\frac{S_{iT,0}}{\widehat{\sigma}^2_{iT,0}}-\mu\right), \end{aligned} \tag{A.26}$$

where $S_{iT,1}$ and $\widehat{\sigma}^2_{iT,1}$ ($S_{iT,0}$ and $\widehat{\sigma}^2_{iT,0}$) stand for S_{iT} and $\widehat{\sigma}^2_{iT}$ evaluated under $\sigma_{v_i^2} > 0$ ($\sigma_{v_i^2} = 0$). We have already shown that as N_2, $T \to \infty$ with $\frac{N_2}{T} \to 0$,

$$\frac{1}{\omega\sqrt{N_2}}\sum_{i=1}^{N_2}\left(\frac{S_{iT,0}}{\widehat{\sigma}^2_{iT,0}}-\mu\right) \Rightarrow N(0,1).$$

Next, notice that for $i \le N_1$, $\widehat{\sigma}^2_{iT,1}$ is the sample mean of a squared $I(1)$ variate and so is $O_p(T)$. It has been shown elsewhere (*e.g.* KPSS) that for $i \le N_1$, $S_{iT,1}$ is $O_p(T^2)$. Therefore,

$$G_{iT} = \frac{S_{iT,1}}{\widehat{\sigma}^2_{iT,1}} - \mu = O_p(T).$$

Under the condition that $\frac{N_1}{N} \to \delta > 0$ $\left(\frac{N_2}{N} \to 1-\delta \ge 0\right)$, it is easily seen that under the alternative hypothesis (2.4),

$$\text{plim}_A \tau_{NT} = \sqrt{\delta} \times \lim_A \sqrt{N_1} \times \text{plim}_A \frac{\sum_{i=1}^{N_1} G_{iT}}{N_1} + \sqrt{1-\delta} \times N(0,1) = \infty,$$

where plim_A and $\lim_A$ denote plim and lim as N, $T \to \infty$, $\frac{N}{T} \to 0$ and $\frac{N_1}{N} \to \delta$, respectively. ■

A.2 Proof of Theorem 3.2

For sake of convenience we first prove (A.5) for the AR(1) case; that is, u_{it}'s in (2.2) follow stationary AR(1) processes with heterogenous coefficients $|\rho_i| < 1$,

$$u_{it} = \rho_i u_{it-1} + \varepsilon_{it}, \tag{A.27}$$

where the ε_{it}'s are *iid* normal variates with zero means and finite heterogenous variances, $\sigma^2_{\varepsilon_i}$. We will get back to the proof for the general AR(p) case at the end of this subsection. Now, it is sufficient to show that (A.18) and (A.25) hold for this case because the rest of the proof of (A.5) is not changed by the process governing the y_{it}.

A.2.1 Proof of (A.18)

It is now sufficient to show that the mean and variance of $\frac{S_{i\infty}-S_{iT}}{\sigma_i^2}$ are bounded functions of T alone and are $O(T^{-1})$ and $o(1)$ in T, respectively and that $\frac{S_{iT}}{\sigma_i^2}$ possesses all moments in the AR(1) case. Then (A.15) to (A.18) may be reapplied intact as before.

(A.7) and (A.8) now become

$$T^2 S_{iT} = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})\,\mathbf{S}\,(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})', \tag{A.28}$$

$$\mathbf{S} = \Omega^{\frac{1}{2}}\mathbf{MPP'M}\Omega^{\frac{1}{2}}, \tag{A.29}$$

where Ω is the $T\times T$ covariance matrix of the vector $(u_{i1}, u_{i2}, ..., u_{iT})'$. Following similar arguments to before, we write $\mathbf{S}$ in its canonical form as $\mathbf{E}\Lambda\mathbf{E}'$ where $\Lambda = diag(\lambda_j)$ are the eigenvalues of $\mathbf{S}$, and $\mathbf{E}$ contains the corresponding orthonormal eigenvectors, gives

$$T^2 S_{iT} = \sum_{t=1}^{T} \lambda_t \varepsilon_{it}^{*\,2},$$

where $(\varepsilon_{i1}^*, \varepsilon_{i2}^*, ..., \varepsilon_{iT}^*) = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})\,\mathbf{E}$ is a $T \times 1$ vector of independent $N(0, \sigma_i^2)$ variates. Using Minkowski's inequality we have

$$\begin{aligned}\left[E\left(\sum_{t=1}^{T}\lambda_t\varepsilon_{it}^{*\,2}\right)^r\right]^{\frac{1}{r}} &\leq \sum_{t=1}^{T}\left\{E\left[\left(\lambda_t\varepsilon_{it}^{*\,2}\right)^r\right]\right\}^{\frac{1}{r}} = \sum_{t=1}^{T}\lambda_t\left[E\left(\varepsilon_{it}^{*\,2r}\right)\right]^{\frac{1}{r}} \\ &= 2\sigma_i^2(2r!)^{\frac{1}{2r}}\sum_{t=1}^{T}\lambda_t = 2(2r!)^{\frac{1}{2r}}E\left(T^2S_{iT}\right),\end{aligned}$$

where the last equality follows from the fact that $\sigma_i^2\sum_{t=1}^{T}\lambda_t = E(T^2S_{iT})$. We shall now compute this last quantity. It is easily established that

$$s_{it} = \frac{1}{1-\rho_i}s_{it}(0) - \frac{\rho_i}{1-\rho_i}\widehat{e}_{it}, \tag{A.30}$$

where $s_{iT}(0) = \sum_{t=1}^{T}\sum_{j=1}^{t}\widehat{\varepsilon}_{ij}$ is constructed from partial sums of de-meaned white noise variates. Squaring (A.30), summing, taking expectations and using (A.14) gives

$$\begin{aligned}E\left(T^2S_{iT}\right) &= E\left[\frac{1}{(1-\rho_i)^2}\sum_{t=1}^{T}s_{it}^2(0) + \left(\frac{\rho_i}{1-\rho_i}\right)^2\sum_{t=1}^{T}\widehat{e}_{it}^2 - \frac{2\rho_i}{(1-\rho_i)^2}\sum_{t=1}^{T}s_{it}(0)\widehat{e}_{it}\right] \\ &= \frac{T^2\sigma_{\varepsilon_i}^2}{6(1-\rho_i)^2} + \frac{\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}\left(\frac{-9T^2-5T+3}{6T}\right) + \\ &\quad \left(\frac{\rho_i}{1-\rho_i}\right)^2\frac{\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}\left[T - \frac{1-2\rho+\rho^{T+1}+T(1-\rho)}{T(1-\rho)^2}\right] \\ &\quad -\frac{2\rho_i}{(1-\rho_i)^2}\sum_{t=1}^{T}s_{it}(0)\widehat{e}_{it}.\end{aligned}$$

It is easy to show that

$$\sum_{t=1}^{T}s_{it}(0)\widehat{e}_{it} = \sum_{t=1}^{T}\widehat{\varepsilon}_{it}\sum_{j=t}^{T}\widehat{e}_{ij}, \tag{A.31}$$

$$\widehat{e}_{ij} = \frac{-\rho_i}{1-\rho_i}\left(\widehat{e}_{ij} - \widehat{e}_{ij-1}\right) + \frac{1}{1-\rho_i}\widehat{\varepsilon}_{ij},$$

where $\widehat{\varepsilon}_{ij} = \varepsilon_{ij} - T^{-1}\sum_{t=1}^{T}\varepsilon_{it}$. Summing $\widehat{e}_{ij}$ from $j = t$ to T and using this in (A.31) gives

$$E\left[\sum_{t=1}^{T}s_{it}(0)\widehat{e}_{it}\right] = \frac{\rho_i}{1-\rho_i}E\left(\sum_{t=1}^{T}\widehat{\varepsilon}_{it}\widehat{e}_{it-1}\right) - \frac{\rho_i}{1-\rho_i}E\left(\widehat{e}_{iT}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}\right) + \frac{\rho_i}{1-\rho_i}E\left(\sum_{t=1}^{T}\widehat{\varepsilon}_{it}s_{it}(0)\right). \tag{A.32}$$

The last expectation term in (A.32) may be written as $(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})\,\mathbf{MP'M}\,(\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})'$, where $\mathbf{M}$ and $\mathbf{P}$ are as defined above. The expectation of this term is therefore just

$$\sigma_{\varepsilon_i}^2 tr\left(\mathbf{MP'M}\right) = \sigma_{\varepsilon_i}^2 tr\left(\mathbf{MP'}\right) = \sigma_{\varepsilon_i}^2 \times \frac{T+1}{2}.$$

Although the first two terms in (A.32) can be readily calculated, anticipating more complex AR(p) processes below, we obtain upper bounds for their absolute value instead. Applying Minkowski's inequality to the first term we see that its absolute value is bounded by an $O(T)$ function of T, $\sigma_{\varepsilon_i}^2$ and ρ_i. Applying Hölder's inequality to the second term we see that its absolute value is bounded by an $O(T)$ function of T, $\sigma_{\varepsilon_i}^2$ and ρ_i. Hence the first term in (A.31) is T^2 times one-sixth of the long run variance of y_{it}. The preceding paragraph establishes that all other terms are bounded functions of T, $\sigma_{\varepsilon_i}^2$ and ρ_i, and are at most $O(T)$ in T. Therefore, dividing (A.31) by $\sigma_i^2 T^2$, where $\sigma_i^2 = \frac{\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}$ is the long run variance of y_{it}, rearranging slightly and taking absolute values we have

$$\left| E\left(\frac{S_{iT}}{\sigma_i^2}\right) - \frac{1}{6} \right| = \left| E\left(\frac{S_{iT} - S_{i\infty}}{\sigma_i^2}\right) \right| \leq \frac{Q_T(\sigma_i^2, \rho_i)}{T}, \tag{A.33}$$

where Q_T has an upper bound that is a function of σ_i^2 and ρ_i and which is $O(1)$ in T. Defining $Q_T^* = \max_{\sigma_i^2 \in \Theta_\sigma, \rho_i \in \Theta_\rho} Q_T(\sigma_i^2, \rho_i)$ we have

$$\left| E\left(\frac{S_{iT} - S_{i\infty}}{\sigma_i^2}\right) \right| \leq \frac{Q_T^*}{T} = \frac{O(1)}{T}, \tag{A.34}$$

which proves the desired result. Finally, using (A.33) and (A.34) in (A.30) establishes that the rth moment of S_{iT} exists.

A.2.2 Proof of (A.25)

We first need to re-express B_{NT}. Using $\sigma_i^2 = \frac{\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}$ and $\hat{\sigma}_{iT}^2 = \frac{\hat{\sigma}_{\varepsilon_i}^2}{(1-\hat{\rho}_i)^2}$, then we obtain

$$\frac{\hat{\sigma}_{iT}^2 - \sigma_i^2}{\hat{\sigma}_{iT}^2} = \left[1 - \frac{(1-\hat{\rho}_i)^2 \sigma_{\varepsilon_i}^2}{(1-\rho_i)^2 \hat{\sigma}_{\varepsilon_{iT}}^2}\right].$$

Using a first order Taylor series expansion for $(1-\hat{\rho}_i)^2$ around ρ_i,

$$(1-\hat{\rho}_i)^2 = (1-\rho_i)^2 - 2\left(1-\rho_i^*\right)\left(\hat{\rho}_i - \rho_i\right),$$

where ρ_i^* lies between $\hat{\rho}_i$ and ρ_i, we have

$$\frac{\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2} = \frac{1}{\hat{\sigma}_{\varepsilon_{iT}}^2}\left[\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right) - \frac{2(1-\rho_i^*)\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}\left(\hat{\rho}_i - \rho_i\right)\right]. \tag{A.35}$$

Substituting (A.35) in the individual component of B_{NT}, we have

$$\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2} = \left(\frac{S_{iT}}{\sigma_i^2}\frac{1}{\hat{\sigma}_{\varepsilon_{iT}}^2}\right) \times \left[\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right) - \frac{2(1-\rho_i^*)\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}\left(\hat{\rho}_i - \rho_i\right)\right]. \tag{A.36}$$

Regarding the second moment and applying the Hölder's inequality we have

$$\begin{aligned} E\left[\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right]^2 &= \frac{1}{T}E\left\{\left(\frac{S_{iT}}{\sigma_i^2}\frac{1}{\hat{\sigma}_{\varepsilon_{iT}}^2}\right)\left[\sqrt{T}\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right) - \frac{2(1-\rho_i^*)\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}\sqrt{T}\left(\hat{\rho}_i - \rho_i\right)\right]\right\} \\ &\leq \frac{1}{T}\sqrt{E\left(\frac{S_{iT}}{\sigma_i^2}\frac{1}{\hat{\sigma}_{\varepsilon_{iT}}^2}\right)^4 E\left[\sqrt{T}\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right) - \frac{2(1-\rho_i^*)\sigma_{\varepsilon_i}^2}{(1-\rho_i)^2}\sqrt{T}\left(\hat{\rho}_i - \rho_i\right)\right]^4}. \end{aligned}$$

The proof is then in five parts: (a) proof of the boundedness of the 8th moment of $\frac{S_{iT}}{\sigma_i^2}$; (b) proof of the boundedness of the 8th moment of $\frac{1}{\hat{\sigma}^2_{\varepsilon_i T}}$; (c) proof of the boundedness of the 4th moment of $\sqrt{T}(\hat{\sigma}^2_{\varepsilon_i T} - \sigma^2_{\varepsilon_i})$; (d) proof of the boundedness of the 4th moment of $\sqrt{T}(\hat{\rho}_i - \rho_i)$;[4] and (e) finally using (a)-(d) to prove (A.25) for the current case.

(a) Proof that the eighth moment of $\frac{S_{iT}}{\sigma_i^2}$ is bounded. The existence of the rth moment of $\frac{S_{iT}}{\sigma_i^2}$ for arbitrary r was established above.

(b) Proof that the eighth moment of $\frac{1}{\hat{\sigma}^2_{\varepsilon_i T}}$ is bounded. We prove the result for even T. The corresponding proof for odd T requires only minor changes to what follows and is omitted for the sake of brevity. Define the data vectors $\mathbf{y}_i' = (y_{i1}, ..., y_{iT})$, $\mathbf{y}_{i,-1}' = (y_{i0}, ..., y_{iT-1})$ and let $\mathbf{i}_k$ denote a $k \times 1$ vector of units. It is a standard result of regression theory that

$$\frac{1}{\hat{\sigma}^2_{\varepsilon_i T}} = T(\mathbf{X}'\mathbf{X})^{-1}_{11}, \tag{A.37}$$

where $(\mathbf{X}'\mathbf{X})^{-1}_{11}$ is the $(1,1)$ element of the inverse of the data product matrix $(\mathbf{X}'\mathbf{X})^{-1}$where $\mathbf{X} = [\mathbf{y}_i, \mathbf{y}_{i,-1}, \mathbf{i}_T]$. Unfortunately, little may be said about the moments of $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ *directly*. Therefore our objective is to find a "smaller" matrix than $\mathbf{X}'\mathbf{X}$ (smaller in the sense that it $\mathbf{X}'\mathbf{X}$ exceeds it by a positive semi-definite (psd) matrix) whose inverse has clear stochastic properties. In what follows the notation $\mathbf{A} \preceq (\succeq)\mathbf{B}$ implies that $\mathbf{A}$ exceeds $\mathbf{B}$ by a psd (nsd) matrix.

We may write $\mathbf{X}'\mathbf{X}$ as

$$\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{X}, \tag{A.38}$$

where $\mathbf{R}$ is a $T \times T$ band matrix with $-\rho_i$ on its super diagonal and units on its main diagonal. It follows that

$$\mathbf{X}'\mathbf{R} = \begin{bmatrix} y_{i1} & \varepsilon_{i2} + \alpha_i & \varepsilon_{i3} + \alpha_i & \cdots & \varepsilon_{iT} + \alpha_i \\ y_{i0} & \varepsilon_{i1} + \alpha_i & \varepsilon_{i2} + \alpha_i & \cdots & \varepsilon_{iT-1} + \alpha_i \\ 1 & 1-\rho_i & 1-\rho_i & \cdots & 1-\rho_i \end{bmatrix} = \mathbf{X}^{*\prime}.$$

Writing $\mathbf{R}'\mathbf{R}$ in its canonical form, (A.38) becomes

$$\mathbf{X}'\mathbf{X} = \mathbf{X}^{*\prime}\mathbf{E}'\Lambda^{-1}\mathbf{E}\mathbf{X}^* = \mathbf{X}^{*\prime}\mathbf{E}'\left(\Lambda^{-1} - \lambda^{-1}_{\max}\mathbf{I}_T\right)\mathbf{E}\mathbf{X}^* + \lambda^{-1}_{\max}\mathbf{X}^{*\prime}\mathbf{X} \succeq \lambda^{-1}_{\max}\mathbf{X}^{*\prime}\mathbf{X}^*,$$

where Λ is the diagonal matrix of ordered eigenvalues of $\mathbf{R}'\mathbf{R}$ whose largest element is $\lambda_{\max}$ and $\mathbf{E}$ is the corresponding matrix of (orthonormal) eigenvectors. Let $\mathbf{X}^*_e$ ($\mathbf{X}^*_o$) denote the $3 \times \frac{T}{2}$ matrix that contains the even numbered (odd numbered) elements of $\mathbf{X}^*$. It follows that

$$\mathbf{X}'\mathbf{X} \succeq \lambda^{-1}_{\max}\mathbf{X}^{*\prime}\mathbf{X}^* = \lambda^{-1}_{\max}\mathbf{X}^{*\prime}_o\mathbf{X}^*_o + \lambda^{-1}_{\max}\mathbf{X}^{*\prime}_e\mathbf{X}^*_e \succeq \lambda^{-1}_{\max}\mathbf{X}^{*\prime}_e\mathbf{X}^*_e. \tag{A.39}$$

Inverting the leftmost and rightmost matrices in (A.39) gives

$$(\mathbf{X}'\mathbf{X})^{-1} \preceq \lambda_{\max}(\mathbf{X}^{*\prime}_e\mathbf{X}^*_e)^{-1}. \tag{A.40}$$

Referring back to the definition in (A.37) and using (A.40) therein gives

$$\frac{1}{\hat{\sigma}^2_{\varepsilon_i T}} = T(\mathbf{X}'\mathbf{X})^{-1}_{11} \leq T\lambda_{\max}(\mathbf{X}^{*\prime}_e\mathbf{X}^*_e)^{-1}_{11}, \tag{A.41}$$

[4]The kth moment of $(1-\rho_i^*)$ is always bounded, if the kth moment of $\sqrt{T}\left(\hat{\rho}_i - \rho_i\right)$ is bounded, since $1-\rho_i^* = (1-\rho_i) + (\rho_i - \rho_i^*)$.

where the subscript $_{11}$ denotes the $(1,1)$th element of the respective matrix and where the *scalar* inequality in (A.41) follows from the fact that if matrix $\mathbf{A}$ exceeds matrix $\mathbf{B}$ by a psd matrix then the diagonal elements of $\mathbf{A}$ (weakly) exceed those of $\mathbf{B}$. The three columns of $\mathbf{X}_e^*$ are $(\varepsilon_{i2}+\alpha_i, \varepsilon_{i4}+\alpha_i, ..., \varepsilon_{iT}+\alpha_i)'$, $(\varepsilon_{i1}+\alpha_i, \varepsilon_{i3}+\alpha_i, ..., \varepsilon_{iT-1}+\alpha_i)'$ and $(1-\rho_i, 1-\rho_i, ..., 1-\rho_i)'$ so that $(\mathbf{X}_e^{*\prime}\mathbf{X}_e^*)_{11}^{-1}$ is equivalent to the residual sum of squares obtained by regressing $(\varepsilon_{i2}, \varepsilon_{i4}, ..., \varepsilon_{iT})'$ on $(\varepsilon_{i1}, \varepsilon_{i3}, ..., \varepsilon_{iT-1})'$ and constants. This quantity may be equivalently written as

$$(\mathbf{X}_e^{*\prime}\mathbf{X}_e^*)_{11}^{-1} = (\mathbf{X}_e'\mathbf{M}\mathbf{X}_e)_{11}^{-1}, \tag{A.42}$$

where $\mathbf{X}_e$ is a $\frac{T}{2}\times 2$ matrix containing $(\varepsilon_{i2}, \varepsilon_{i4}, ..., \varepsilon_{iT})'$ and $(\varepsilon_{i1}, \varepsilon_{i3}, ..., \varepsilon_{iT-1})'$, and $\mathbf{M} = \mathbf{I}_{\frac{T}{2}} - \mathbf{i}_{\frac{T}{2}}\left(\mathbf{i}_{\frac{T}{2}}'\mathbf{i}_{\frac{T}{2}}\right)^{-1}\mathbf{i}_{\frac{T}{2}}'$ is the idempotent matrix. Therefore, we have shown [*e.g.* Press (1972, p. 113)]

$$(\mathbf{X}_e^{*\prime}\mathbf{X}_e^*)_{11}^{-1} = IW\left(2, \frac{T}{2}-1\right)_{11}. \tag{A.43}$$

Combining (A.40) - (A.43) gives

$$\frac{1}{\hat{\sigma}_{\varepsilon_i T}^2} \leq T\lambda_{\max}(\mathbf{X}_c'\mathbf{X}_c)_{11}^{-1} = T\lambda_{\max} IW\left(2, \frac{T}{2}-1\right)_{11}. \tag{A.44}$$

Now, it is easily verified that $\mathbf{R}'\mathbf{R}$ is a band matrix with $-\rho_i$ on its sub and super diagonals and the elements $\left(1, 1+\rho_i^2, 1+\rho_i^2, ..., 1+\rho_i^2\right)$ on its main diagonal. We may rewrite its canonical form as

$$\mathbf{E}\mathbf{R}'\mathbf{R}\mathbf{E}' = \Lambda. \tag{A.45}$$

Expanding the $(1,1)$th element of (A.45) gives an explicit form for $\lambda_{\max}$ as

$$\begin{aligned}\lambda_{\max} &= \left(\mathbf{E}\mathbf{R}'\mathbf{R}\mathbf{E}'\right)_{11} = \left(1+\rho_i^2\right)\sum_{t=1}^{T} m_{1t}^2 - \rho_i^2\sum_{t=1}^{T} m_{1t}^2 - 2\rho_i\sum_{t=2}^{T} m_{1t}m_{1t-1} \\ &= 1 - 2\rho_i\sum_{t=2}^{T} m_{1t}m_{1t-1} \leq 1 + 2\rho_i,\end{aligned}$$

where m_{1t} is the $(1,t)$th element of $\mathbf{E}$ and the equality and inequality in the second line of (A.46) follow by virtue of the unit length of the eigenvectors and by application of Schwartz's inequality respectively. (A.46) and the strict positivity of $\lambda_{\max}$ shows that its eighth power is a bounded function of ρ_i, T and $\sigma_{\varepsilon_i}^2$ $\forall T$, and $\forall \rho_i \in \Theta_\rho$ and $\sigma_{\varepsilon_i}^2 \in \Theta_\sigma$, where Θ_ρ and Θ_σ are compact sets. By direct integration using the formula for the *pdf* of a diagonal element of an $IW\left(2, \frac{T}{2}-1\right)$ matrix (*e.g.* Press (1972, p. 111)) it is easy to show that T^8 times the eighth moment of such an element is bounded and $O(1)$ in T. These facts prove the desired result.

(c) Proof that the 4th moment of $\sqrt{T}\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right)$ is bounded. Using standard regression theory we have

$$\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2 = \frac{1}{T}\left(\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i - T\sigma_{\varepsilon_i}^2\right),$$

where $\boldsymbol{\varepsilon}_i' = (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})$ and $\mathbf{M}_i = \mathbf{I}_T - \mathbf{Z}_i\left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'$ with $\mathbf{Z}_i = (\mathbf{i}_T, \mathbf{y}_{i-1})$, $\mathbf{i}_T = (1, 1, ..., 1)'$ and $\mathbf{y}_{i,-1} = \left(y_{i0}, y_{i1}, ..., y_{iT-1}\right)'$. Using the fact that

$$\boldsymbol{\varepsilon}_i'\mathbf{M}_i\boldsymbol{\varepsilon}_i = \sum_{t=1}^{T}\varepsilon_{it}^2 - \boldsymbol{\varepsilon}_i'\mathbf{P}_i\boldsymbol{\varepsilon}_i \geq 0,$$

where $\mathbf{P}_i = \mathbf{Z}_i\left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i$, it is easily seen that

$$\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2 \leq \frac{1}{T}\left(\sum_{t=1}^{T}\varepsilon_{it}^2 - T\sigma_{\varepsilon_i}^2\right) = \frac{\sigma_{\varepsilon_i}^2}{T}\left(\sum_{t=1}^{T}\frac{\varepsilon_{it}^2}{\sigma_{\varepsilon_i}^2} - T\right).$$

Notice that $\sum_{t=1}^{T} \frac{\varepsilon_{it}^2}{\sigma_{\varepsilon_i}^2} = \chi_T^2$, where χ_T^2 stands for a χ^2 variate with T degrees of freedom. Since $E\left(\chi_T^2\right) = T$, hence it is well known that $E\left(\chi_T^2 - T\right)^4 = 12T\,(T+1)$ [*e.g.* Poirier (1995, p. 101)]. Therefore,

$$E\left[\sqrt{T}\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right)\right]^4 \le E\left[\frac{\sigma_{\varepsilon_i}^2}{\sqrt{T}}\left(\sum_{t=1}^{T}\frac{\varepsilon_{it}^2}{\sigma_{\varepsilon_i}^2} - T\right)\right]^4 = \frac{\sigma_{\varepsilon_i}^8}{T^2} \times 12T\,(T+1)\,,$$

and the last term here is clearly bounded.

(d) Proof that the fourth moment of $\sqrt{T}\,(\hat{\rho}_i - \rho_i)$ is bounded. Because $\hat{\rho}_i$ involves filtered data using the estimator of the MA(1) parameter, $\hat{\theta}_i$, we first prove a general result concerning the expected value of positive functions of $\hat{\theta}_i$. For the proof here we further assume that an estimator of θ has the following properties: (i) $\hat{\theta}_i$ is a T-consistent estimator of θ_i; (ii) $\hat{\theta}_i$ is uniquely determined by $\{y_{i1}, y_{i2}, ..., y_{iT}\}$; (iii) $\hat{\theta}_i$ is constrained to take one of r discrete values defined in the set $\digamma_r = \left\{0, \frac{1}{r}, \frac{2}{r}, ..., \frac{r-1}{r}, 1\right\}$ where r is a finite integer greater than 1; (iv) under the alternative hypothesis (2.4), plim $\hat{\theta}_i \ne 1$ for $i = 1, ..., N_1$; (v) the 64th moment of $T\left(1 - \hat{\theta}_i\right)$ exists. Note that any estimator of θ can be made to satisfy (iii) by a suitable numerical rounding procedure or "discretisation" of the underlying (continuous) values. An example of an estimator satisfying all of these assumptions is a discretised version of the estimator given in Snell (1999). A discretised version of Choi's (1992) GLS estimator of $\hat{\theta}_i$ would also satisfy (i) to (iv). The property (v) may well be true for the latter, but a proof of this is beyond scope of the current paper.

Since we have assumed $\hat{\theta}_i$ to be identified, the sample space of raw data $\mathbf{y}_i = (y_{i0}, y_{i1}, ..., y_{iT})'$, which we denote by Y, can be divided into $r+1$ subspaces Y_j, $j = 0, 1, ..., r$, such that

$$\hat{\theta}_i\left(\mathbf{y}_i\right) = \frac{j}{r} \quad iff \quad \mathbf{y}_i \in Y_j, \tag{A.46}$$

where $\hat{\theta}_i(\mathbf{y}_i)$ denotes the mapping from the data vector $\mathbf{y}_i$ to $\hat{\theta}_i$. Consider the weakly positive function $g\left(\hat{\theta}_i(\mathbf{y}_i), \mathbf{y}_i\right)$. We may write the expected value of $g(\cdot)$ as

$$E\left[g\left(\hat{\theta}_i(\mathbf{y}_i), \mathbf{y}_i\right)\right] = \int_Y g\left(\hat{\theta}_i(\mathbf{y}_i), \mathbf{y}_i\right) f(\mathbf{y}_i) d\mathbf{y}_i = \sum_{j=0}^{r} \int_{Y_j} g\left(\frac{j}{r}, \mathbf{y}_i\right) f(\mathbf{y}_i) d\mathbf{y}_i, \tag{A.47}$$

where $f(\mathbf{y}_i)$ is the multivariate normal probability distribution function and with a slight abuse of notation, $\int_Y \cdots d\mathbf{y}_i$ and $\int_{Y_j} \cdots d\mathbf{y}_i$ denote integration with respect to the $T+1$ variates $y_{i0}, y_{i1}, ..., y_{iT}$ over $Y \in R^{T+1}$ and over $Y_j \in R^{T+1}$, respectively. The positivity of $g(\cdot)$ and normality of the variates mean that we can expand further to get

$$\sum_{j=0}^{r} \int_{Y_j} g\left(\frac{j}{r}, \mathbf{y}_i\right) f(\mathbf{y}_i) d\mathbf{y}_i \le \sum_{j=0}^{r} \int_{Y} g\left(\frac{j}{r}, \mathbf{y}_i\right) f(\mathbf{y}_i) d\mathbf{y}_i = \sum_{j=0}^{r} E\left[g\left(\frac{j}{r}, \mathbf{y}_i\right) f(\mathbf{y}_i)\right]. \tag{A.48}$$

(A.47) and (A.48) imply that the expected value of any (positive) expression $g\left(\hat{\theta}_i(\mathbf{y}_i), \mathbf{y}_i\right)$ will be bounded if the expected values of $g\left(\frac{j}{r}, \mathbf{y}_i\right)$ are bounded for all $j = 0, 1, ..., r$.

Noting that under $\sigma_{v_i^2} = 0$,

$$\Delta\tilde{y}_{it}^f = \rho_i \Delta\tilde{y}_{it-1}^f + \Delta\varepsilon_{it}^f, \tag{A.49}$$

we may write $\sqrt{T}\,(\hat{\rho}_i - \rho_i)$ as

$$\sqrt{T}(\hat{\rho}_i - \rho_i) = \left(\frac{1}{T}\sum_{t=1}^{T} \Delta\tilde{y}_{it-1}^{f\ 2}\right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Delta\tilde{y}_{it-1}^f \Delta\varepsilon_{it}^f = \frac{A}{B},\ say. \tag{A.50}$$

Applying Hölders inequality to the right most expression in (A.50) we require boundedness of the 8th moments of A and $\frac{1}{B}$, respectively. Consider A first. Using (A.49) and the identity,

$$\Delta x_t^f \equiv x_t - \left(1-\hat{\theta}_i\right) x_{t-1}^f, \tag{A.51}$$

for any variable x_t, we now have

$$\sqrt{T}(\hat{\rho}_i - \rho_i) = \left\{ \frac{1}{T} \sum_{t=1}^{T} \left[\tilde{y}_{it-1} - \left(1-\hat{\theta}_i\right) \tilde{y}_{it-2}^f \right]^2 \right\}^{-1}$$

$$\times \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\tilde{y}_{it-1}\varepsilon_{it} - \left(1-\hat{\theta}_i\right)\left(\tilde{y}_{it-2}^f \varepsilon_{it} + \tilde{y}_{it-1}\varepsilon_{it-1}^f\right) + \left(1-\hat{\theta}_i\right)^2 \tilde{y}_{it-2}^f \varepsilon_{it-1}^f \right] = \frac{A}{B}, \tag{A.52}$$

where we have used the fact that $\sum_{t=1}^{T} \tilde{a}_t \tilde{b}_t = \sum_{t=1}^{T} \tilde{a}_t b_t$ to express ε_{it} as a levels variate rather than a deviation from mean variate. We may decompose A as

$$A = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{it} - T\left(1-\hat{\theta}_i\right) \frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} \left(\tilde{y}_{it-2}^f \varepsilon_{it} + \tilde{y}_{it-1}\varepsilon_{it-1}^f\right) + \left[T\left(1-\hat{\theta}_i\right)\right]^2 \frac{1}{T^{\frac{5}{2}}} \sum_{t=1}^{T} \tilde{y}_{it-2}^f \varepsilon_{it-1}^f. \tag{A.53}$$

Applying Minkowski's and Hölder's inequality to (A.53), we have

$$\begin{aligned} \left[E\left(A^8\right)\right]^{\frac{1}{8}} \leq & \left[E\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{it}\right)^8\right]^{\frac{1}{8}} + \left\{E\left[T\left(1-\hat{\theta}_i\right)\right]^{16}\right\}^{\frac{1}{16}} \left\{E\left[\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} \left(\tilde{y}_{it-2}^f \varepsilon_{it} + \tilde{y}_{it-1}\varepsilon_{it-1}^f\right)\right]^{16}\right\}^{\frac{1}{16}} \\ & + \left\{E\left[\left(T\left(1-\hat{\theta}_i\right)\right)^2\right]^{16} E\left(\frac{1}{T^{\frac{5}{2}}} \sum_{t=1}^{T} \tilde{y}_{it-2}^f \varepsilon_{it-1}^f\right)^{16}\right\}^{\frac{1}{16}}. \end{aligned}$$

For boundedness of the 8th moment of A, we now require that the 8th moment of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{it}$, the 16th moments of $\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} \left(\tilde{y}_{it-2}^f \varepsilon_{it} + \tilde{y}_{it-1}\varepsilon_{it-1}^f\right)$ and $\frac{1}{T^{\frac{5}{2}}} \sum_{t=1}^{T} \tilde{y}_{it-2}^f \varepsilon_{it-1}^f$, and the 32nd moment of $T\left(1-\hat{\theta}_i\right)$ are all bounded. The last of these holds by assumption. It follows that to establish the boundedness of other expressions it is sufficient to show that the expectations of the other three terms for fixed $\hat{\theta}_i = 1$ and for fixed $\hat{\theta}_i \in [0, 1)$ are respectively bounded.

Consider first the case for $\hat{\theta}_i = 1$. In this special case (A.49) reduces to

$$\tilde{y}_{it} = \rho_i \tilde{y}_{it-1} + \varepsilon_{it},$$

and therefore, we require existence of 8th moment of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{it}$, and 16th moments of $\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} \varepsilon_{it} \sum_{j=1}^{t} \tilde{y}_{ij-2}$, $\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} \tilde{y}_{it-1} \sum_{j=1}^{t} \varepsilon_{ij-1}$, and $\frac{1}{T^{\frac{5}{2}}} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} \tilde{y}_{ij-2}\right)\left(\sum_{j=1}^{t} \varepsilon_{ij-1}\right)$. Using Minkowski's inequality on the first of these terms we have

$$\begin{aligned} \left[E\left(\sum_{t=1}^{T} \tilde{y}_{it-1}\varepsilon_{it}\right)^8\right]^{\frac{1}{8}} &= \left[E\left(\sum_{t=1}^{T} u_{it-1}\varepsilon_{it} - \overline{y}_{it-1} \sum_{t=1}^{T} \varepsilon_{it}\right)^8\right]^{\frac{1}{8}} \qquad \text{(A.54)} \\ &\leq \left[E\left(\sum_{t=1}^{T} u_{it-1}\varepsilon_{it}\right)^8\right]^{\frac{1}{8}} + \left[E\left(\overline{y}_{it-1} \sum_{t=1}^{T} \varepsilon_{it}\right)^8\right]^{\frac{1}{8}}. \end{aligned}$$

Applying Holder's inequality to the second expectation term on the far right gives

$$E\left(\overline{y}_{it-1}\sum_{t=1}^{T}\varepsilon_{it}\right)^{8} \leq \sqrt{E\left(\overline{y}_{it-1}^{16}\right)E\left[\left(\sum_{t=1}^{T}\varepsilon_{it}\right)^{16}\right]} = O(1),$$

where the last equality follows from the fact that $\overline{y}_{it-1} \sim N\left\{0, O(T^{-1})\right\}$, and $\sum_{t=1}^{T}\varepsilon_{it} \sim N\left\{0, O(T)\right\}$. Boundedness of $\frac{1}{\sqrt{T}}\left[E\left(\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{it}\right)^{16}\right]^{\frac{1}{16}}$ therefore rests on boundedness of $\frac{1}{\sqrt{T}}\left[E\left(\sum_{t=1}^{T}u_{it-1}\varepsilon_{it}\right)^{8}\right]^{\frac{1}{8}}$. Expanding and applying Minkowski's inequality we get[5]

$$\left[E\left(\sum_{t=1}^{T}u_{it-1}\varepsilon_{it}\right)^{8}\right]^{\frac{1}{8}} = \left[E\left(\sum_{j=1}^{\infty}\rho_i^j\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}\right)^{8}\right]^{\frac{1}{8}}$$

$$\leq \left(\sum_{j=1}^{\infty}\rho_i^j\right)\left[E\left(\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}\right)^{8}\right]^{\frac{1}{8}} \leq \left(\sum_{j=1}^{\infty}\rho_i^j\right)\max_j\left[E\left(\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}\right)^{8}\right]^{\frac{1}{8}}.$$

Adopt the shorthand $\sum_{t=1}^{T}a_t$ for $\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}$ for some arbitrary (fixed) j, and consider the expansion of its eighth power. We divide the terms in the expansion into three types. The first type are $O(T^4)$ in number and take the form $a_i^{k_1}a_j^{k_2}a_k^{k_3}a_l^{k_4}$, where k_j, $j = 1, 2, 3, 4$, are weakly positive integers summing to 8 with $k_1 = 5$ and where $i \neq j \neq k \neq l$. By applying Hölders inequality to $E\left(a_i^{k_1}a_j^{k_2}a_k^{k_3}a_l^{k_4}\right)$ we find that these terms are each $O(1)$ in magnitude. There are $O(T^4)$ of such terms so that their total contribution to $E\left(\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}\right)^{8}$ is of order T^4. The second set of terms take the form of either $a_i^4a_ja_ka_la_m$ or $a_i^3a_ja_ka_la_ma_n$ or $a_j^2a_ka_la_ma_na_oa_p$ or $a_ia_ja_ka_la_ma_na_oa_p$, and at first sight would appear to be in excess of $O(T^4)$ in number. However, each of these contains a term $\varepsilon_{it-j-1}\varepsilon_{it}$ such that either ε_{it-j-1} or ε_{it} in this term appears nowhere else in the multiple. As a result their expected value factors into either $E(\varepsilon_{it-j-1}) \times E(\textit{other term})$ or $E(\varepsilon_{it}) \times E(\textit{other term})$ and hence are all equal to zero. The third type of terms have the form $a_i^3a_j^2a_ka_la_m$ or $a_i^2a_j^2a_ka_la_ma_n$, which again appear to be more numerous than $O(T^4)$. However, using the previous argument plus the fact that third moment of normal variates is zero, we find that only the second of this type has non-zero expectation. A necessary condition for $E\left(a_i^2a_j^2a_ka_la_ma_n\right)$ to be non-zero is that either $i = \min(i, j, k, l, m, n)$ and $j = \max(i, j, k, l, m, n)$ or $i = \max(i, j, k, l, m, n)$ and $j = \min(i, j, k, l, m, n)$. These conditions restrict the number of non-zero expectation terms of this type to at most $O(T^4)$.[6] As with the first type of terms $E\left(a_i^2a_j^2a_ka_la_ma_n\right)$ is $O(1)$. This proves that $E\left(\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}\right)^{8}$ is of $O(T^4)$ regardless of j so that $E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it-j-1}\varepsilon_{it}\right)^{8}$ is bounded.

We now deal with $\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{t}\tilde{y}_{ij-2}$. We may write

$$\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{t}\tilde{y}_{ij-2} = \frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{t}\tilde{y}_{ij} - \frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\varepsilon_{it}\tilde{y}_{it-1} - \frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\varepsilon_{it}\tilde{y}_{it}. \tag{A.55}$$

Applying Minkowski's inequality (e.g. in the manner of (A.54)) and noting that the second term on the right of (A.55) has just been shown to be bounded, it is sufficient to show that the first and third terms

[5] Note that $\sum_{j=1}^{\infty}\rho_i^j$ is the sum of the MA coefficients in the Wold form for y_{it} and so is bounded. It is also bounded for any stationary process so the details of this section of the proof would not change for the AR(p) case below.

[6] In fact there are far fewer than $O(T^4)$ terms because there is a further necessary condition for the expectation being non-zero. The necessary condition given in the text is suffucient for our purposes.

on the right of (A.55) have bounded 16th moments. A straightforward and by now familiar application of Minkowski's inequality shows that this holds true for the second term on the right. The first term may be written as

$$\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{t}\tilde{y}_{ij}=\frac{1}{T^{\frac{3}{2}}}\left(\varepsilon_{i1},\ldots,\varepsilon_{iT}\right)\Omega^{\frac{1}{2}}\mathbf{PM}\left(\varepsilon_{i1},\ldots,\varepsilon_{iT}\right)',$$

where $\Omega^{\frac{1}{2}}$, $\mathbf{P}$ and $\mathbf{M}$ are as defined earlier. Reapplying the argument in (A.28) to (A.30), the 16th moment of the above is bounded if $\sum_{t=1}^{T}\lambda_t=\sigma_{\varepsilon_i}^2 tr\left(\Omega^{\frac{1}{2}}\mathbf{PM}\right)=E\left(\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{t}\tilde{y}_{ij}\right)$ is $O(T^{\frac{3}{2}})$ or less. Using the fact that the columns and rows of the covariance matrix of a stationary variable are summable it is easy to show that $tr\left(\Omega^{\frac{1}{2}}\mathbf{PM}\right)=O(T)$, and hence that the 16th moment of $T^{-\frac{3}{2}}\sum_{t=1}^{T}\varepsilon_{it}\sum_{j=1}^{t}\tilde{y}_{ij}$ is bounded.

A nearly identical argument applies to $T^{-\frac{3}{2}}\sum_{t=1}^{T}\tilde{y}_{it-1}\sum_{j=1}^{t}\varepsilon_{ij-1}$.

Finally, to establish the boundedness of the 16th moment of $T^{-\frac{5}{2}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\tilde{y}_{ij-2}\right)\left(\sum_{j=1}^{t}\varepsilon_{ij-1}\right)$ we may use the same arguments for the third term in (A.55) with minor modifications. Write it as a quadratic form in $(\varepsilon_{i1},\ldots,\varepsilon_{iT})$ plus other terms to which Minkowski's inequality may be applied to show boundedness and then analyse the trace of the matrix in the quadratic form. The latter is $tr\left(\Omega^{\frac{1}{2}}\mathbf{MPP}'\right)$ and by tedious expansion (and again using the summability of stationary autocovariances) is easily shown to be $O(T^2)$. Hence the 16th moment of $T^{-\frac{5}{2}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\tilde{y}_{ij-2}\right)\left(\sum_{j=1}^{t}\varepsilon_{ij-1}\right)$ is bounded.

Next, consider the case for $0\leq\hat{\theta}_i<1$. By successive application of Minkowski's and Hölder's inequality to $\frac{1}{T}\sum_{t=1}^{T}\tilde{y}_{it-2}^{f}\varepsilon_{it}$, we may write

$$\left[E\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\tilde{y}_{it-2}^{f}\varepsilon_{it}\right)^{16}\right]^{\frac{1}{16}}=\frac{1}{T^{\frac{3}{2}}}\left[E\left(\sum_{t=1}^{T}\tilde{y}_{it-2}^{f}\varepsilon_{it}\right)^{16}\right]^{\frac{1}{16}}$$

$$\leq\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\left[E\left(\tilde{y}_{it-2}^{f}\varepsilon_{it}\right)^{16}\right]^{\frac{1}{16}}\leq\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\left[E\left(\tilde{y}_{it-2}^{f}\right)^{32}\right]^{\frac{1}{16}}\times\left[E(\varepsilon_{it})^{32}\right]^{\frac{1}{16}}=\frac{O(1)}{\sqrt{T}},\tag{A.56}$$

where the last equality follows from the fact that $\tilde{y}_{it-2}^{f}$ and ε_{it} are zero-mean normal variates with variances that are finite functions of T, ρ_i and $\sigma_{\varepsilon i}^2$ and that are $O(1)$ in T. (A.56) shows that $T^{-\frac{3}{2}}\sum_{t=1}^{T}\tilde{y}_{it-2}^{f}\varepsilon_{it}$ has sixteenth moment that is bounded for $0\leq\hat{\theta}_i<1$. Exactly the same arguments may be applied to show boundedness of $\left[E\left(\frac{1}{T^{\frac{3}{2}}}\sum_{t=1}^{T}\tilde{y}_{it-1}\varepsilon_{it-1}^{f}\right)^{16}\right]^{\frac{1}{16}}$ and $\left[E\left(\frac{1}{T^{\frac{5}{2}}}\sum_{t=1}^{T}\tilde{y}_{it-2}^{f}\varepsilon_{it-1}^{f}\right)^{16}\right]^{\frac{1}{16}}$, and for reasons of brevity we do not repeat them here.

Turning now to the 8th moment of $\frac{1}{B}$ in (A.52), we may write

$$\frac{1}{B}=T\left\{\sum_{t=1}^{T}\left[\tilde{y}_{it-1}-\left(1-\hat{\theta}_i\right)\tilde{y}_{it-2}^{f}\right]^2\right\}^{-1}.\tag{A.57}$$

Consider the case for $\hat{\theta}_i=1$ first. In this case $\frac{1}{B}$ simply reduces to

$$T\left(\sum_{t=1}^{T}\tilde{y}_{it-1}^2\right)^{-1}=T\left(\mathbf{X}'\mathbf{X}\right)_{11}^{-1},\tag{A.58}$$

where $\mathbf{X}$ is now the $T\times 2$ matrix whose columns are $(y_1,...,y_T)'$ and $(1,1,...,1)'$. The template of the proof given in part (b) above may be adapted with only minor and obvious amendments to show that the 8th moment of the expression in (A.58) is $O(1)$ in T.

Next, for the case of $0 \leq \hat{\theta}_i < 1$ we may apply a similar argument to that given above for $T\left(\sum_{t=1}^{T}\tilde{y}_{it-1}^2\right)^{-1}$ to show that the term in braces on the right of (A.57) is bounded by an $IW\left(1, \frac{T-1}{2}\right)$ variate. Following the development in (A.38) we may write

$$\left[\sum_{t=1}^{T}\left(\tilde{y}_{it-1} - (1-\hat{\theta}_i)\tilde{y}_{it-2}^f\right)^2\right]^{-1} = \left[\mathbf{X}'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{X}\right]_{11}^{-1}$$

where now $\mathbf{R} = \mathbf{A}(\theta_i)\mathbf{B}(\rho_i)$, $\mathbf{A}(\theta_i)$ and $\mathbf{B}(\rho_i)$ have unity on their main diagonals and $-\theta_i$ and $-\rho_i$ respectively on their superdiagonals and other notations are the same as above. Under the current definition of $\mathbf{R}$,

$$\mathbf{X}'\mathbf{R} = \left\{\tilde{y}_{iT-1} - (1-\hat{\theta}_i)\tilde{y}_{iT-2}^f, \Delta\varepsilon_{iT-1}, \Delta\varepsilon_{iT-1}, ..., \Delta\varepsilon_{i0}\right\}.$$

Following previous arguments,

$$\left[\mathbf{X}'\mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{X}\right]_{11}^{-1} \preceq \lambda_{\max}\left(\mathbf{X}'\mathbf{R}\mathbf{R}'\mathbf{X}\right)_{11}^{-1}$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}'\mathbf{R}$. Below in the proof of the AR(p) case we show that the largest eigenvalue of the covariance matrix of a $T \times 1$ vector of observations on a stationary variate is bounded. It is readily established that $\mathbf{R}'\mathbf{R}$ is such a covariance matrix so this bounding argument applies here also. Denoting $\mathbf{X}_e$ as the $2 \times \frac{T}{2}$ matrix containing the even numbered elements of $\mathbf{X}'\mathbf{R}$ and using previous arguments gives

$$\lambda_{\max}\left(\mathbf{X}'\mathbf{R}\mathbf{R}'\mathbf{X}\right)_{11}^{-1} \preceq \lambda_{\max}\left(\mathbf{X}_e\mathbf{X}_e\right)_{11}^{-1}.$$

The variate on the right is an $IW\left(1, \frac{T-1}{2}\right)$. T^k times the kth moment of such a variate is a bounded constant and is $O(1)$ in T, so that the 8th moment of $\frac{1}{B}$ is bounded. Hence we have shown that the 8th moment of $\frac{1}{B}$ is bounded for $0 \leq \hat{\theta}_i \leq 1$. This completes the proof that the 4th moment of $\sqrt{T}(\hat{\rho}_i - \rho_i)$ is bounded.

(e) Proof of (A.5) for the AR(1) case. To prove (A.5) for the AR(1) case we follow exactly the same procedure as for the simple case and establish a fixed upper bound for the second moment of d_{iT}. As before, write d_{iT} as

$$d_{iT} = \frac{(S_{i\infty} - S_{iT})}{\sigma_i^2} + \frac{1}{\sqrt{T}}\left(\frac{S_{iT}}{\sigma_i^2}\right)\left[\frac{\sqrt{T}\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2}\right], \tag{A.59}$$

and show that its second moment is bounded and is $O(T^{-1})$ in T. As before we deal with the first and second terms in (A.59), separately. We have shown above that the mean and variance of $\frac{S_{i\infty}-S_{iT}}{\sigma_i^2}$ are bounded functions of T, $\sigma_{\varepsilon_i}^2$ and ρ_i and are $O(T^{-1})$ and $o(1)$ in T respectively. Denoting $\sqrt{T}$ times the mean of this term by $M_T\left(\sigma_{\varepsilon_i}^2, \rho_i\right)$ and the variance of this term by $V_T\left(\sigma_{\varepsilon_i}^2, \rho_i\right)$ we may write

$$M_T^* = \max_{\sigma_{\varepsilon_i}^2 \in \Theta_\sigma, \rho_i \in \Theta_\rho} M_T\left(\sigma_{\varepsilon_i}^2, \rho_i\right) = O\left(T^{-\frac{1}{2}}\right);\ V_T^* = \max_{\sigma_{\varepsilon_i}^2 \in \Theta_\sigma, \rho_i \in \Theta_\rho} V_T\left(\sigma_{\varepsilon_i}^2, \rho_i\right) = o(1). \tag{A.60}$$

Recalling the definition of D_{NT} given by (A.4) and using the independence of panels, then

$$Var(A_{NT}) = \frac{1}{N}\sum_{i=1}^{N}Var\left(\frac{S_{i\infty} - S_{iT}}{\sigma_i^2}\right) \leq \frac{1}{N}\sum_{i=1}^{N}V_T^* = V_T^* = o(1), \tag{A.61}$$

$$E(A_{NT}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left(\frac{S_{i\infty} - S_{iT}}{\sigma_i^2}\right) \leq \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{M_T^*}{\sqrt{T}} = \sqrt{\frac{N}{T}}M_T^* = \sqrt{\frac{N}{T}} \times O(T^{-\frac{1}{2}}), \tag{A.62}$$

where the last equalities follow from (A.60). This shows that

$$\lim_{N,T\to\infty,\frac{N}{T}\to 0} E(A_{NT}) = 0 \quad and \quad \lim_{N,T\to\infty,\frac{N}{T}\to 0} Var(A_{NT}) = 0, \tag{A.63}$$

which establishes via Chebyshev's inequality that

$$\text{plim}_{N,T\to\infty,\frac{N}{T}\to 0} A_{NT} = 0. \tag{A.64}$$

We now show that the same result holds for B_{NT} also. As (A.4) shows, B_{NT} is the normalised sum of the terms

$$\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2} = \frac{1}{\sqrt{T}}\left(\frac{S_{iT}}{\sigma_i^2}\right)\left(\frac{1}{\hat{\sigma}_{iT}^2}\right)\left[\sqrt{T}\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)\right] = \frac{1}{\sqrt{T}}E_{iT}F_{iT}G_{iT},\ say. \tag{A.65}$$

Using the results obtained in subsections (a) and (d) above and Hölder's inequality on $E_{iT}F_{iT}G_{iT}$ establishes the boundedness of the second moment of $\frac{S_{iT}}{\sigma_i^2}\frac{\left(\hat{\sigma}_{iT}^2-\sigma_i^2\right)}{\hat{\sigma}_{iT}^2}$. Each upper bound in (a) - (d) above was established for *any* admissible $T, \sigma_{\varepsilon_i}^2$ and ρ_i (*i.e.* for $T > T_L$ and for any $\sigma_{\varepsilon_i}^2 \in \Theta_\sigma$ and $\rho_i \in \Theta_\rho$). Furthermore, it is readily established by inspection that the second moment of $E_{iT}F_{iT}G_{iT}$ is *solely* a function of T, $\sigma_{\varepsilon_i}^2$ and ρ_i. Proceeding as before then we have

$$E\left(E_{iT}F_{iT}G_{iT}\right)^2 = C_T(\sigma_{\varepsilon_i}^2,\rho_i) \leq C_T^* < \infty, \tag{A.66}$$

where

$$C_T^* = \max_{\sigma_{\varepsilon_i}^2\in\Theta_\sigma,\rho_i\in\Theta_\rho} C_T(\sigma_{\varepsilon_i}^2,\rho_i)\ for\ T > T_L. \tag{A.67}$$

From Jensen's inequality it follows that

$$E\left(E_{iT}F_{iT}G_{iT}\right) \leq \sqrt{C_T^*} < \infty. \tag{A.68}$$

Using (A.66) - (A.68), the mean and variance of B_{NT} satisfy

$$E(B_{NT}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{T}}E\left(E_{iT}F_{iT}G_{iT}\right) \leq \sqrt{\frac{N}{T}}\sqrt{C_T^*}, \tag{A.69}$$

$$Var(B_{NT}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}Var\left(E_{iT}F_{iT}G_{iT}\right) \leq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}E\left(E_{iT}F_{iT}G_{iT}\right)^2 \leq \frac{C_T^*}{T}. \tag{A.70}$$

These results demonstrate that

$$\lim_{N,T\to\infty,\frac{N}{T}\to 0} E(B_{NT}) = 0; \quad \lim_{N,T\to\infty,\frac{N}{T}\to 0} Var(B_{NT}) = 0, \tag{A.71}$$

which in turn via Chebyshev's inequality imply that

$$\text{plim}_{N,T\to\infty,\frac{N}{T}\to 0} B_{NT} = 0. \tag{A.72}$$

From the definition of D_{NT} given in (A.4), (A.64) and (A.72) establish (A.5) for the AR(1) case.

A.2.3 Extension to the $AR(p)$ case

The proof is readily extended to the general $AR(p)$ case, following near identical arguments to those used above. To avoid repetition, we only provide arguments and equations where they differ from the previous case.

For proof of (A.18): As before, boundedness of higher moments of S_{iT} rests on boundedness of its mean. The mean could be computed directly as before by extending (A.30) to (A.32). However, this becomes long winded for the $AR(p)$ case. Instead we note that $E(S_{iT}) = tr\left(\Omega^{\frac{1}{2}}\mathbf{MPP}'\mathbf{M}\Omega^{\frac{1}{2}}\right) = tr\left(\Omega\mathbf{MPP}'\mathbf{M}\right)$]. Tedious and uninformative expansion (available on request) shows that this trace is the sum of $O(T^2)$, $O(T)$ and $O(1)$ terms. It follows that $E(S_{iT})$ is bounded and hence so are higher moments. The remaining results follow without change.

For proof of (A.25): Since the long run variance in the $AR(p)$ case is now $\sigma_i^2 = \frac{\sigma_{\varepsilon_i}^2}{\rho_i(1)^2}$, hence the analogue of (A.35) for the current case is

$$\frac{\left(\hat{\sigma}_{iT}^2 - \sigma_i^2\right)}{\hat{\sigma}_{iT}^2} = \frac{1}{\hat{\sigma}_{\varepsilon_i T}^2}\left\{\left(\hat{\sigma}_{\varepsilon_i T}^2 - \sigma_{\varepsilon_i}^2\right) - 2\sum_{j=1}^{p}\frac{(1-\rho_{ij}^*)\sigma_{\varepsilon_i}^2}{(1-\rho_{ij})^2}\left(\hat{\rho}_{ij} - \rho_{ij}\right)\right\}, \tag{A.73}$$

where ρ_{ij}^* lies between $\hat{\rho}_{ij}$ and ρ_{ij}. The remaining arguments carry over here.

For subsection (a): The existence of the rth moment of $\frac{S_{iT}}{\sigma_i^2}$ for arbitrary r was established above.

For subsection (b): We now have

$$\frac{1}{\hat{\sigma}_{\varepsilon_i T}^2} = T(\mathbf{X}'\mathbf{X})_{11}^{-1}, \tag{A.74}$$

where $\mathbf{X} = (\mathbf{y}_i, \mathbf{y}_{i,-1}, ..., \mathbf{y}_{i,-p}, \mathbf{i}_T)$ is a $T \times (p+2)$ data matrix. The transformation matrix $\mathbf{R}$ now has units on the main diagonal and $-\rho_{ij}$ on the jth superdiagonal. Then the arguments proceed as before up to (A.38). At that point we selected the even numbered elements of the data in $\mathbf{X}^*$ whereas here we select the $(p+1)$th numbered elements. For example, when $p = 10$, we would select the 11th, 22nd, 33rd, ..., rows of $\mathbf{X}^*$ and place them in the $(p+1) \times \frac{T}{p+1}$ data matrix again denoted by $\mathbf{X}_e^*$. The remaining arguments go through up to (A.42) and (A.43) except that the limiting random variable is now $IW\left(p+1, \frac{T-1}{p+1}\right)$ rather than $IW\left(2, \frac{T-1}{2}\right)$. Finally, the maximum eigenvalue expression (A.46) becomes

$$\begin{aligned}\lambda_{\max} &= \sum_{j=1}^{T} m_{1j}^2 r_{jj} + 2\sum_{k=1}^{T}\sum_{j=i}^{T} m_{1k}m_{1j}r_{kj} = m_{11}^2\left(r_{11} - r_D\right) + r_D + 2\sum_{k=1}^{T}\sum_{j=i}^{T} m_{1k}m_{1j}r_{kj} \\ &\leq m_{11}^2\left(r_{11} - r_D\right) + r_D + 2\sqrt{\left(\sum_{i=1}^{T}\sum_{j=i}^{T} m_{1i}m_{1j}r_{ij}\right)^2} \leq m_{11}^2\left(r_{11} - r_D\right) + r_D + 2r_{\max} < \infty,\end{aligned}$$

where m_{1j} is the $(1, j)$th element of the matrix of the corresponding orthonormal eigenvectors, $\mathbf{E}$, r_{kj} denotes the (k, j)th element of $\mathbf{R}'\mathbf{R}$, $r_{\max}$ is the largest (in absolute value) off-diagonal element of $\mathbf{R}$, and r_D denotes the identical elements r_{kk} $(k > 1)$. In deriving the above equalities and inequalities, we have used (i) unit length of the eigenvectors, (ii) positivity of the eigenvalue, (iii) Schwartz's inequality and (iv) the 2-summability of the vector of AR coefficients $(\rho_{i1}, ..., \rho_{ip})$, respectively.

For subsection (c): The arguments carry over directly to the current case except that $\mathbf{Z}_i$ is now the $T \times (p+1)$ matrix $(\mathbf{y}_{i-1}, ..., \mathbf{y}_{i-p}, \mathbf{i}_T)$.

For subsection (d): Now $\boldsymbol{\rho}_i$ is the $p \times 1$ vector $\boldsymbol{\rho}_i = (\rho_{i1}, \rho_{i2}, ..., \rho_{ip})'$ so that

$$\sqrt{T}\left(\hat{\boldsymbol{\rho}}_i - \boldsymbol{\rho}_i\right) = \left[\left(\frac{\Delta\mathbf{Z}_i^{f'}\Delta\mathbf{Z}_i^f}{T}\right)^{-1}\frac{\Delta\mathbf{Z}_i^{f'}\Delta\boldsymbol{\varepsilon}_i^f}{\sqrt{T}}\right]_p = \left(\mathbf{B}^{-1}\mathbf{A}\right)_p, \ say, \tag{A.75}$$

where the subscript p denotes the first p elements of the respective matrix and $\mathbf{Z}_i = [\tilde{\mathbf{y}}_{i,-1}, \tilde{\mathbf{y}}_{i,-2}, ..., \tilde{\mathbf{y}}_{i,-p}]$ is the $T \times (p+1)$ data matrix containing T observations on the p lagged demeaned series. We show that all $(p+1)$ elements of $\mathbf{B}^{-1}\mathbf{A}$ have the necessary bounded moments.

As before we may apply Hölder's and Minkowski's inequality to establish that the boundedness of the eighth moment of each element of $\mathbf{B}^{-1}\mathbf{A}$ rests on the boundedness of the sixteenth moments of each element of $\mathbf{B}^{-1}$ and $\mathbf{A}$, respectively. We start with $\mathbf{B}^{-1}$ and again we show the required boundedness results for $\hat{\theta}_i = 1$ and for $0 \leq \hat{\theta}_i < 1$ separately.

For $\hat{\theta}_i = 1$, $\mathbf{B}^{-1}$ becomes $T\left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}$ and the analysis in section (b) may be directly reapplied here with the following minor changes: (i) $\mathbf{X}$ now becomes $(\mathbf{Z}_i, \mathbf{i}_T)$, (ii) $\mathbf{R}$ is re-defined for the $AR(p)$ case as given above, every $(p+1)$th observation is collated in $\mathbf{X}_e^*$ rather than every 2nd, and (iii) the subscript '$_{11}$' after a matrix now denotes the upper $(p+1) \times (p+1)$ elements rather than the $(1,1)$th element. With these amendments, the analysis in (b) may be reapplied to get a bounding matrix for $T\left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}$ in the manner of (A.43). This bounding matrix is now $T\lambda_{\max}$ times an $IW\left(p+1, \frac{T-1}{p+1}\right)$ variate. The eighth moments of all elements of T times the IW matrix are bounded and $\lambda_{\max}$, the largest eigenvalue of $\mathbf{R}'\mathbf{R}$ has been shown above to be bounded so that the 8th moments of each element in the matrix $T\left(\mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}$is also bounded.

Only minor amendments need be made to apply the preceding arguments to the case of $0 \leq \widehat{\theta} < 1$. The transformation matrix $\mathbf{R}$ is now the product of two $T \times T$ matrices, *i.e.* $\mathbf{R} = \mathbf{R}_1\mathbf{R}_2$ where $\mathbf{R}_1$ has ones on its diagonal and $-\widehat{\theta}_i$ on its superdiagonal and $\mathbf{R}_2$ has ones on its main diagonal and $-\rho_j$ on its jth superdiagonal. (Apart from the initial terms which are subsequently dropped, the first matrix transforms $\Delta\mathbf{Z}_i^f$ to $\Delta\mathbf{Z}_i$ and the second transforms $\Delta\mathbf{Z}_i$ to the first differenced white noise data $\Delta\boldsymbol{\varepsilon}_i$). The limiting IW variate becomes $IW\left(p+1, \frac{T-1}{p+1}\right)$. The only remaining difficulty is showing that $\lambda_{\max}$ (the largest eigenvalue of $\mathbf{R}_2'\mathbf{R}_1'\mathbf{R}_1\mathbf{R}_2$) is bounded for this case. To show this, consider the $p \times 1$ random vector $\mathbf{v} = \mathbf{R}_2'\mathbf{R}_1'\boldsymbol{\omega}$, where $E(\boldsymbol{\omega}) = \mathbf{0}$ and $E(\boldsymbol{\omega}\boldsymbol{\omega}') = \mathbf{I}_T$. Clearly, $\mathbf{v}$ is a stationary process with covariance matrix $\mathbf{R}_2'\mathbf{R}_1'\mathbf{R}_1\mathbf{R}_2$ and bounded off-diagonal elements. We may therefore re-apply the argument at the end of subsection (b) to the current case.

On expansion of the matrix $\mathbf{A}$, it can be seen that each of its elements has the same form as the term in (A.53). Proof of boundedness of each of these elements therefore proceeds along identical lines as the $AR(1)$ case. The only added complication is the more general form of the covariance matrix of the y_{it}. However, the only characteristic of this matrix used in the previous proofs was the summability of its columns and rows. This property holds for the rows and columns of the covariance matrix of T observations on *any* stationary variate.

For subsection (e): The arguments in subsection (e) carry over directly to the current case.

Finally, the consistency of the test can be proved along similar lines to the proof of Theorem 3.2.

References

Baltagi, B.H.; Kao, C. Panel Time Series Models: A Survey. Advances in Econometrics (forthcoming) **2002**.

Banerjee, A. Panel Data Unit Roots and Cointegration: An Overview. Unpublished manuscript, Oxford University **1999**.

Billingsley, P. *Probability and Measure*; John Wiley and Sons: New York, **1979**.

Chang, Y. Nonlinear IV Unit Root Test in Panels with Cross-sectional Dependency. Unpublished manuscript, Rice University, **2000**.

Choi, I. Asymptotic Distributions of the Conditional Nonlinear Least Squares Estimates for Noninvertible MA(1) Processes. Unpublished manuscript, Ohio State University, **1992**.

Choi, I. Unit Root Tests for Panel Data. Journal of International Money and Finance **2001**, 20, 249-272.

Hadri, K. Testing for Stationarity in Heterogeneous Panel Data. Econometrics Journal **2000**, 3, 148-161.

Im, K.S.; Pesaran, M.H.; Shin, Y. Testing for Unit Roots in Heterogenous Panel. Journal of Econometrics (forthcoming) **2002**.

Kao, C. Spurious Regression and Residual-based Tests for Cointegration in Panel Data. Journal of Econometrics **1999**, 90, 1-44.

Kauppi, H. Panel Data Limit Theory and Asymptotic Analysis of a Panel Regression with Near Integrated Regressors. Unpublished manuscript, University of Helsinki, **2000**.

Kiviet, J.; Phillips, G. Bias Assessment and Reduction in Linear Error-correction Models. Journal of Econometrics **1994**, 63, 215-243.

Kwiatkowski, D.; Phillips, P.C.B.; Schmidt, P.; Shin, Y. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Are Nonstationary. Journal of Econometrics **1992**, 54, 159-178.

Levin, A.; Lin, C.F. Unit Root Tests in Panel Data: Asymptotic and Finite-sample Properties. Unpublished manuscript, University of California, San Diego, **1993**.

Leybourne, S.; McCabe, P.B.M.; On Estimating an ARMA Model with an MA Unit Root. Econometric Theory **1998**, 14, 326-338.

Maddala, G.S.; Wu, S. A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test. Oxford Bulletin of Economics and Statistics **1999**, 61, 631-652.

McCoskey, S.; Kao, C. A Residual-based Tests of the Null of Cointegration in Panel Data. Econometric Reviews, **1998**, 17, 57-84.

Pedroni, P. Critical Values for Cointegration Test in Heterogeneous Panels with Multiple Regressors. Oxford Bulletin of Economics and Statistics **1999**, 61, 653-670.

Pesaran, M.H.; Smith, R.P. Estimating Long-run Relationships from Dynamic Heterogeneous Panels. Journal of Econometrics **1995**, 68, 79-113.

Pesaran M.H.; Shin, Y.; Smith, R.P. Pooled mean group estimation of dynamic heterogenous panels. Journal of the American Statistical Association **1999**, 94, 621-634.

Phillips, P.C.B.; Moon, R.H. Linear Regression Limit Theory and Non-stationary Panel Data. Econometrica **1999**, 67, 1157-1112.

Phillips, P.C.B.; Sul, D. Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. Unpublished manuscript, Yale University, **2002**.

Poirier, D. *Intermediate Statistics and Econometrics*. MIT Press: Cambridge, **1995**.

Press, S.J. *Applied Multivariate Analysis*. Holt, Rinehart and Winston: New York, **1972**.

Quah, D. Exploiting Cross-section Variations for Unit Root Inference in Dynamic Data. Economics Letters **1994**, 44, 9-19.

Smith, R.P. Estimation and Inference with Non-stationary Panel Time Series Data. Unpublished manuscript, Birkbeck College, **2000**.

Snell, A. Testing for r versus r-1 Cointegrating Vectors. Journal of Econometrics **1999**, 88, 151-191.

Summers, R.; Heston, A. The Penn World Table (Mark 5): An expanded set of international comparisons 1950-1988. Quarterly Journal of Economics **1991**, 106, 327-368.