A Note on the Pooling of Individual PANIC Unit Root Tests*

Joakim Westerlund† Rolf Larsson‡

September 19, 2007

Abstract

One of the most cited studies in recent years within the field of nonstationary panel data analysis is that of Bai and Ng (2004), in which the authors propose PANIC, a new framework for analyzing the nonstationarity of panels with idiosyncratic and common components. This paper shows that their results are not sharp enough to ensure PANIC as an asymptotically valid framework for constricting pooled panel unit root tests.

JEL Classification: C21; C22; C23.

Keywords: Panel Unit Root Test; Pooling; Common Factor; Cross-Sectional Dependence.

1 Introduction

Consider the observed variable X_{it}, where $t = 1, ..., T$ and $i = 1, ..., N$ indexes the time series and cross-sectional units, respectively. The starting point of PANIC is to decompose X_{it} into two components, one that is common across i and one that is idiosyncratic. In this note, we

*Previous versions of this paper were presented at Lund University, the University of Maastricht and the University of Cambridge. The authors would like to thank seminar participants, and in particular Jushan Bai, David Edgerton, Serena Ng, Jean-Pierre Urbain, Hashem Pesaran, Vanessa Smith and Takashi Yamagata for many valuable comments and suggestions. The first author would also like to thank the Maastricht Research School of Economics of Technology and Organizations for its hospitality during a visit at the Department of Quantitative Economics at the University of Maastricht, where a part of this paper was written. Financial support from the Jan Wallander and Tom Hedelius Foundation under research grant numbers P2005-0117:1 and W2006-0068:1 is gratefully acknowledged. The usual disclaimer applies.

†Corresponding author: Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. Telephone: +46 46 222 8670; Fax: +46 46 222 4118; E-mail address: joakim.westerlund@nek.lu.se.

‡Department of Information Science, Division of Statistics, Uppsala University, P. O. Box 513, S-751 20 Uppsala, Sweden. E-mail address: rolf.larsson@dis.uu.se.

consider the simple setup with an intercept only, in which case X_{it} may be written as

$$X_{it} = c_i + \lambda_i' F_t + e_{it} = c_i + \sum_{j=1}^{r} \lambda_{ji} F_{jt} + e_{it}, \tag{1}$$

where the common factor F_{jt} and loading λ_{ji} represent the common component of X_{it}, while e_{it} represents the idiosyncratic component. These are assumed to be generated as

$$F_{jt} = \phi_j F_{jt-1} + u_{jt} \quad \text{and} \quad e_{it} = \rho_i e_{it-1} + \epsilon_{it}, \tag{2}$$

where we assume for simplicity that u_{jt} and ϵ_{it} are uncorrelated across t. In this setup, the idiosyncratic component e_{it} has a unit root if $\rho_i = 1$ and it is stationary if $\rho_i < 1$. Similarly, if some of the ϕ_j parameters are equal to one, then X_{it} has as many common stochastic trends as the number of unit roots in F_t.

The objective of PANIC is to determine the number of common stochastic trends and test if $\rho_i = 1$ when F_t and e_{it} are estimated using the method of principal components. The problem is that if e_{it} is nonstationary, then this method cannot be applied to X_{it} because it will render the resulting estimate of λ_i inconsistent. Bai and Ng (2004) therefore suggest applying the principal components method to x_{it}, the first difference of X_{it}, rather than to X_{it} itself. To appreciate the point of this, note that x_{it} can be written as

$$x_{it} = \lambda_i' f_t + z_{it}, \tag{3}$$

where f_t and z_{it} are the first differences of F_t and e_{it}, respectively. In contrast to (1), all the components of this equation are stationary, which means that consistent estimates $\widehat{\lambda}_i$, $\widehat{f}_t$ and $\widehat{z}_{it} = x_{it} - \widehat{\lambda}_i' \widehat{f}_t$ of λ_i, f_t and z_{it} can be obtained. Estimates $\widehat{F}_t$ and $\widehat{e}_{it}$ of F_t and e_{it} can then be obtained by simply recumulating $\widehat{f}_t$ and $\widehat{z}_{it}$.

The idea behind PANIC is to test whether $\rho_i = 1$ by subjecting $\widehat{e}_{it}$ to any conventional unit root test, such as the classical Dickey and Fuller (1979) test, which can be written as $DF^c_{\widehat{e}}(i) = \frac{\widehat{U}_i}{\sqrt{\widehat{V}_i}}$ where $\widehat{U}_i$ and $\widehat{V}_i$ are the usual sample moments of $\widehat{e}_{it}$. The justification for testing in this particular way is that $DF^c_{\widehat{e}}(i)$ is asymptotically equivalent to $DF^c_e(i)$, the unit root test based on e_{it}. Similarly, knowing $\widehat{F}_t$ is as good as knowing F_t, in the sense that $DF^c_{\widehat{F}}(i)$ is asymptotically equivalent to $DF^c_F(i)$. This is very convenient as it implies that it is possible to disentangle the sources of potential nonstationarity in X_{it} by separately testing for unit roots in e_{it} and F_t.

Another interesting advantage of PANIC is that $DF^c_{\hat{e}}(i)$ can be used to construct pooled tests for a unit root in e_{it}. The conventional way to construct such tests involves first demeaning the data, and then subjecting each of the demeaned series to a unit root test.[1] If X_{it} is independent across i, the normalized average of these tests converges to a normal variate under the null hypothesis of a unit root. Unfortunately, such tests are generally inappropriate as X_{it} will usually exhibit at least some form of dependence across i. By contrast, pooled tests based on e_{it} are more widely applicable, since they are valid under the more plausible assumption that X_{it} admits to a common factor structure.

Yet another advantage, even in comparison to other studies that also permit for common factors, is that in PANIC the factors need not be stationary. This makes tests based on e_{it} very general indeed, and is probably one of the main reasons why PANIC has become so popular in both applied and theoretical work, see Breitung and Pesaran (2005).

This paper points out a weakness in PANIC that seems to have been largely overlooked in the literature. In particular, it is shown that the theoretical results provided by Bai and Ng (2004) are not enough to ensure that PANIC can be used for constructing asymptotically valid panel tests based on averaging. This is because the order of the error incurred when replacing $DF^c_e(i)$ with $DF^c_{\hat{e}}(i)$ is not sufficiently sharp to ensure that it vanishes as N increases. This paper provides additional results showing that PANIC can in fact be used for pooling purposes.

2 Main results

This section reports our main results using as an example the $DF^c_{\hat{e}}(i)$ statistic, which was also considered by Bai and Ng (2004). However, the results apply to all panel tests that are based on pooling across individual test statistics or their p-values. The data generating process is taken directly from Bai and Ng (2004), and consists of (1) and (2) plus their assumptions A through E. For simplicity, in this paper we also assume that u_{it} and ϵ_{it} are serially uncorrelated, and that ϵ_{it} is normally distributed.[2]

[1] See Breitung and Pesaran (2005) for a recent survey of the existing panel unit root literature.

[2] The first assumption is by no means restrictive, in the sense that violations can be easily accommodated by using any serial correlation corrected test, such as the augmented Dickey and Fuller (1979) test. Although we speculate that the second assumption can be relaxed as well, some of the proofs given in this paper would not go through, and normality is therefore maintained.

The appendix shows that if $\rho_i = 1$ holds for unit i, then

$$DF^c_{\widehat{e}}(i) \;=\; DF^c_e(i) + \mathcal{R}_i \;=\; DF^c_e(i) + O_p\left(\frac{1}{C_{NT}}\right) \;\Rightarrow\; \mathcal{B}_i \quad \text{as} \quad N,\, T \to \infty, \tag{4}$$

where $\mathcal{R}_i$ is a remainder term, $C_{NT} = \min\{\sqrt{T},\, \sqrt{N}\}$ and $\mathcal{B}_i$ is the usual Dickey and Fuller (1979) test distribution. The by far most common way of pooling statistics of this sort is to take the cross-sectional average, $\overline{DF}^c_{\widehat{e}}(N)$ say. Bai and Ng (2004) argues that since $DF^c_{\widehat{e}}(i)$ is asymptotically equivalent to $DF^c_e(i)$ and $\mathcal{B}_i$ is independent across i, it must be true that

$$\sqrt{N}(\overline{DF}^c_{\widehat{e}}(N) - E(\mathcal{B})) \;\Rightarrow\; N(0, \mathrm{var}(\mathcal{B}))$$

and this should hold irrespectively of the relative expansion rate of N and T, as long as they both go to infinity. However, this is not correct, as seen by noting that as $N,\, T \to \infty$

$$\begin{aligned} \sqrt{N}(\overline{DF}^c_{\widehat{e}}(N) - E(\mathcal{B})) \;&=\; \sqrt{N}(\overline{DF}^c_{e}(N) - E(\mathcal{B})) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathcal{R}_i. \\ &\Rightarrow\; N(0, \mathrm{var}(\mathcal{B})) + O_p\left(\frac{\sqrt{N}}{C_{NT}}\right). \end{aligned} \tag{5}$$

Thus, even if we assume that $\frac{N}{T} \to 0$ as $N,\, T \to \infty$, the order of the remainder is still $O_p(1)$. In other words, although still valid on an individual unit level as seen in (4), based on the results provided by Bai and Ng (2004), PANIC does not seem to be a valid approach for pooling tests. The following theorem shows that this suspicion is uncalled for, and that the $O_p(1)$ remainder actually is $o_p(1)$.

Theorem 1. *Under the assumptions given above and the null hypothesis that $\rho_i = 1$ for all i, as $N,\, T \to \infty$ with $\frac{N}{T} \to 0$*

(a) $\overline{DF}^c_{\widehat{e}}(N) \;\to_p\; E(\mathcal{B})$

(b) $\sqrt{N}(\overline{DF}^c_{\widehat{e}}(N) - E(\mathcal{B})) \;\Rightarrow\; N(0, \mathrm{var}(\mathcal{B}))$

A detailed account of the remainder in (5) is provided in the appendix. However, it is instructive to note that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathcal{R}_i \;=\; -\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\frac{e_{iT}A_{iT}}{\sqrt{V_i}} + \frac{1}{\sqrt{N}T^2}\sum_{i=1}^{N}\frac{\sigma^2_{\epsilon i}U_i}{V_i^{3/2}}\sum_{t=2}^{T}e_{it-1}A_{it-1} + O_p\left(\frac{\sqrt{N}}{R_{NT}}\right),$$

where $R_{NT} = \min\{\sqrt{T},\, N\}$, A_{it} is the cumulative sum of $a_{it} = \widehat{\lambda}'_i\widehat{f}_t - \lambda'_i f_t$, and U_i and V_i are $\widehat{U}_i$ and $\widehat{V}_i$ based on e_{it}, respectively. It is further possible to show that if $N,\, T \to \infty$ with

$\frac{N}{T} \to 0$, then this expression has the following limit

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathcal{R}_i \ \to_p \ \phi \left(\lim_{N,T\to\infty} \sqrt{N} E(\mathcal{R}) \right),$$

where $\sqrt{N}E(\mathcal{R})$ tends to zero as N, $T \to \infty$. Our simulation results suggest that ϕ plays a big role in determining the small-sample performance of the pooled test. To see why, note that ϕ can be written as

$$\phi \ = \ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_{\epsilon i}} \sqrt{\lambda_i' \left(\Sigma^{-1} \Gamma \Sigma^{-1} \right) \lambda_i},$$

where Γ and Σ are such that $\frac{1}{N} \sum_{i=1}^{N} \sigma_{\epsilon i}^2 \lambda_i \lambda_i' \to \Gamma$ and $\frac{1}{N} \sum_{i=1}^{N} \lambda_i \lambda_i' \to \Sigma$ as $N \to \infty$, respectively. Thus, since ϕ is essentially an average of N unit specific variance ratios, it is generally positive, which means that $\sqrt{N}(\overline{DF}_{\hat{e}}^c(N) - E(\mathcal{B}))$ is expected to be negative in small samples, as $\sqrt{N}E(\mathcal{R})$ will tend to be negative. Being left-tailed, this is suggestive of an oversized test. We also see that $\sigma_{\epsilon i}^2$ and λ_i can be scaled by a constant without leaving any effect on ϕ. As an extreme example, note that if $\sigma_{\epsilon i}^2$ is equal across i, then Γ simplifies to $\sigma_\epsilon^2 \Sigma$ in which case we get

$$\phi \ = \ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_\epsilon} \sqrt{\lambda_i' \left(\Sigma^{-1} \Gamma \Sigma^{-1} \right) \lambda_i} \ = \ \frac{1}{N} \sum_{i=1}^{N} \sqrt{\lambda_i' (\Sigma^{-1}) \lambda_i}.$$

Thus, if ϵ_{it} is homoskedastic across i then there is no dependence upon σ_ϵ^2. Heteroskedasticity is therefore an important factor in determining the extent of any small-sample bias that comes from replacing $DF_{\hat{e}}^c(i)$ with $DF_e^c(i)$.

3 Simulations

A small-scale simulation was conducted to assess the impact of our asymptotic results in small samples. For that purpose, since our focus lies in examining the null distribution of the test, the data are generated according to (1) and (2) with $\rho_i = 1$ for all i. For simplicity, we further assume that $\phi_j = 0$ for all j, that $r = 2$, that $\lambda_i \sim N(0,1)$ and that $c_i = 1$ for all i. The errors u_{jt} and ϵ_{it} are both assumed to be mean zero and normally distributed with variance one and $\sigma_{\epsilon i}^2 \sim U(0,b)$, respectively. The parameter b determines the degree of the heteroskedasticity of the idiosyncratic component, and is key in the simulations. All computations have been performed in GAUSS using 10,000 replications. The results reported in Table 1 may be summarized as follows.[3]

Firstly, looking at the rightmost column, we see that the simulated remainder is a significant contributor to the variation of $\overline{DF}_{\widehat{e}}(N)$ with a variance share close to 50% in most cases. We also see that this share slowly disappears as N and T grows, which corroborates the asymptotic results.

Secondly, it interesting to see how the centering of the simulated remainder is affected by N and T on the one hand, and by b on the other hand. With b fixed, we see that while the effect of increasing T seem to be small, a larger N pushes the distribution of the remainder, and hence also that of $\overline{DF}_{\widehat{e}}(N)$, to the right, thus making positive outcomes more likely. Hence, since the critical region is in the left tail of the normal distribution, this will make the test more conservative. By contrast, if we fix N and T, and instead let b increase, we see that the remainder tends to the left, causing $\overline{DF}_{\widehat{e}}(N)$ to become oversized. When there is no heteroskedasticity, there is no bias effect.

Finally, note that while the performance of the pooled tests seems to be greatly affected by the parametrization of the data generating process, as expected, the performance of $DF_{\widehat{e}}(i)$ is essentially unaffected.

4 Concluding remarks

In this paper we point out a flaw in the theoretical results provided by Bai and Ng (2004) for their PANIC unit root methodology. The problem lies in the order of the error incurred when replacing the idiosyncratic component e_{it} by its estimated counterpart $\widehat{e}_{it}$, which is not sharp enough to ensure that PANIC can be used for pooling across i. The current paper provides more exact results establishing that PANIC can in fact be used for pooling purposes.

[3] To better isolate the effect of pooling, we have assumed that the true number of factors is known. Also, as in the previous sections, we do not provide any results for the case with serial correlation. Interested readers are referred to the paper of Kapetanios (2007) for some results when the data are serially correlated.

Appendix: Mathematical proofs

In this appendix, we prove Theorem 1. In so doing, we will make use of the fact that the common factor can only be identified up to a scale matrix H, say. Thus, what we will consider here is the rotation HF_t of F_t. As usual, $||A||$ will denote the Euclidean norm $\sqrt{tr(A'A)}$ of the matrix A.

Lemma A.1. As $N, T \to \infty$

$$\text{(a)} \quad \widehat{V}_i = V_i - \sigma_{\epsilon i}^2 \left(\frac{2}{T^2} \sum_{t=2}^{T} e_{it-1} A_{it-1} \right) + O_p\left(\frac{1}{R_{NT}} \right),$$

$$\text{(b)} \quad \widehat{U}_i = U_i - \frac{1}{T} e_{iT} A_{iT} + O_p\left(\frac{1}{R_{NT}} \right).$$

Proof of Lemma A.1.

We begin with (a). By definition

$$\widehat{V}_i = \widehat{\sigma}_{\epsilon i}^2 \frac{1}{T^2} \sum_{t=2}^{T} \widehat{e}_{it-1}^2.$$

Consider $\widehat{\sigma}_{\epsilon i}^2$. Under the null hypothesis that $\rho_i = 1$ for all i

$$\begin{aligned} \widehat{\sigma}_{\epsilon i}^2 &= \frac{1}{T} \sum_{t=1}^{T} \widehat{\epsilon}_{it}^2 = \frac{1}{T} \sum_{t=2}^{T} (\Delta \widehat{e}_{it} - (\widehat{\rho}_i - 1) \widehat{e}_{it-1})^2 \\ &= \frac{1}{T} \sum_{t=2}^{T} (\Delta \widehat{e}_{it})^2 - 2(\widehat{\rho}_i - 1) \left(\frac{1}{T} \sum_{t=2}^{T} \widehat{e}_{it-1} \Delta \widehat{e}_{it} \right) + T(\widehat{\rho}_i - 1)^2 \left(\frac{1}{T^2} \sum_{t=2}^{T} \widehat{e}_{it-1}^2 \right). \quad \text{(A1)} \end{aligned}$$

Consider the first term on the right-hand side of this equation. From the text, we have that the defactored and first differentiated residuals can be written as

$$\Delta \widehat{e}_{it} = x_{it} - \widehat{\lambda}_i' \widehat{f}_t = \Delta e_{it} - (\widehat{\lambda}_i' \widehat{f}_t - \lambda_i' f_t) = \Delta e_{it} - a_{it}. \quad \text{(A2)}$$

If we let $d_i = \widehat{\lambda}_i - (H^{-1})' \lambda_i$ and $v_t = \widehat{f}_t - H f_t$, then a_{it} can be rewritten as

$$a_{it} = \lambda_i' H^{-1} (\widehat{f}_t - H f_t) + (\widehat{\lambda}_i - (H^{-1})' \lambda_i)' \widehat{f}_t = \lambda_i' H^{-1} v_t + d_i' \widehat{f}_t. \quad \text{(A3)}$$

This implies

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2 &= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it}-a_{it})^2 \\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2+\frac{1}{T}\sum_{t=2}^{T}a_{it}^2-\frac{2}{T}\sum_{t=2}^{T}\Delta e_{it}a_{it} \\
&= \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2+I-II, \quad \text{say.} \qquad \text{(A4)}
\end{aligned}$$

By using the same arguments as in Lemma B.1 of Bai and Ng (2004), part I is $O_p(1/C_{NT}^2)$. Part II can be written as

$$\begin{aligned}
II &= 2\lambda_i' H^{-1}\frac{1}{T}\sum_{t=2}^{T}\Delta e_{it}v_t+2d_i'\frac{1}{T}\sum_{t=2}^{T}\Delta e_{it}\widehat{f}_t \\
&= 2d_i'\frac{1}{T}\sum_{t=2}^{T}\Delta e_{it}\widehat{f}_t+O_p\left(\frac{1}{C_{NT}^2}\right) \\
&= 2d_i'\frac{1}{T}\sum_{t=2}^{T}\Delta e_{it}(\widehat{f}_t-Hf_t)+2d_i'H\frac{1}{T}\sum_{t=2}^{T}\Delta e_{it}f_t+O_p\left(\frac{1}{C_{NT}^2}\right),
\end{aligned}$$

where the second equality follows by Lemma B.1 of Bai (2003). By applying $||AB|| \leq ||A||||B||$ and the triangle inequality to the remaining part, we get

$$|II| \leq 2||d_i||\left(\frac{1}{T}\sum_{t=2}^{T}||\Delta e_{it}(\widehat{f}_t-Hf_t)||\right)+2||d_i||||H||\left(\frac{1}{T}\sum_{t=2}^{T}||\Delta e_{it}f_t||\right)+O_p\left(\frac{1}{C_{NT}^2}\right),$$

which, by applying the Cauchy-Schwarz inequality to the first term on the right-hand side, reduces to

$$\begin{aligned}
|II| &\leq 2||d_i||\left(\frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=2}^{T}||\widehat{f}_t-Hf_t||^2\right)^{1/2} \\
&+ 2||d_i||||H||\left(\frac{1}{T}\sum_{t=2}^{T}||\Delta e_{it}f_t||\right)+O_p\left(\frac{1}{C_{NT}^2}\right) \\
&= O_p\left(\frac{1}{R_{NT}}\right)O_p(1)O_p\left(\frac{1}{C_{NT}}\right)+O_p\left(\frac{1}{R_{NT}}\right)O_p\left(\frac{1}{\sqrt{T}}\right)+O_p\left(\frac{1}{C_{NT}^2}\right),
\end{aligned}$$

where $||H|| = O_p(1)$ by construction and $||d_i|| = O_p(1/R_{NT})$ by Lemma 1 (c) of Bai and Ng (2004). Also, from Lemma A.1 of Bai (2003), we have

$$\frac{1}{T}\sum_{t=2}^{T}||\widehat{f}_t-Hf_t||^2 = O_p\left(\frac{1}{C_{NT}^2}\right). \qquad \text{(A5)}$$

This implies that II is $O_p(1/C_{NT}^2)$, which in turn implies

$$\frac{1}{T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2 \;=\; \frac{1}{T}\sum_{t=2}^{T}(\Delta e_{it})^2 + O_p\left(\frac{1}{C_{NT}^2}\right). \tag{A6}$$

Consider next the third term on the right-hand side of (A1), which, since $\widehat{e}_{i1} = 0$ by definition, may be written as

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=2}^{T}\widehat{e}_{it}^2 &= \frac{1}{T^2}\sum_{t=2}^{T}(e_{it} - e_{i1} - A_{it})^2 \\
&= \frac{1}{T^2}\sum_{t=2}^{T}e_{it}^2 + \frac{1}{T}e_{i1}^2 + \frac{1}{T^2}\sum_{t=2}^{T}A_{it}^2 - \frac{2}{T^2}\sum_{t=2}^{T}e_{it}A_{it} \\
&- 2e_{i1}\left(\frac{1}{T^2}\sum_{t=2}^{T}e_{it} - \frac{1}{T^2}\sum_{t=2}^{T}A_{it}\right) \\
&= \frac{1}{T^2}\sum_{t=2}^{T}e_{it}^2 + I + II - III - IV, \quad \text{say.}
\end{aligned} \tag{A7}$$

Part I is obviously $O_p(1/T)$. The next step is to show that II is $O_p(1/C_{NT}^2)$. We begin by using (A3), which implies that $A_{it} = \lambda_i' H^{-1}V_t + d_i'\widehat{F}_t$ where V_t is the cumulative sum of v_t. Thus, by subsequently applying the triangle inequality and then $(a+b)^2 \le 2(a^2+b^2)$, we get

$$\begin{aligned}
|II| &= \frac{1}{T^2}\left|\left|\sum_{t=2}^{T}(\lambda_i' H^{-1}V_t + d_i'\widehat{F}_t)^2\right|\right| \\
&\le 2||\lambda_i' H^{-1}||^2\left(\frac{1}{T^2}\sum_{t=2}^{T}||V_t||^2\right) + 2||d_i||^2\left(\frac{1}{T^2}\sum_{t=2}^{T}||\widehat{F}_t||^2\right) \\
&= O_p\left(\frac{1}{C_{NT}^2}\right) + O_p\left(\frac{1}{R_{NT}^2}\right)O_p(1),
\end{aligned}$$

where we have used equation (A.3) of Bai and Ng (2004), which says that

$$\frac{1}{\sqrt{T}}V_t \;=\; O_p\left(\frac{1}{C_{NT}}\right) \tag{A8}$$

and therefore

$$\frac{1}{T^2}\sum_{t=2}^{T}||V_t||^2 \;=\; \frac{1}{T}\sum_{t=2}^{T}\left(\left|\left|\frac{1}{\sqrt{T}}V_t\right|\right|\right)^2 \;=\; O_p\left(\frac{1}{C_{NT}^2}\right). \tag{A9}$$

Also, Lemma B.2 (i) of Bai and Ng (2004) implies that $\frac{1}{T^2}\sum_{t=2}^{T}||\widehat{F}_t||^2 = O_p(1)$.

Part III can be rewritten as

$$\begin{aligned} III &= 2\lambda_i' H^{-1} \frac{1}{T^2} \sum_{t=2}^{T} e_{it} V_t + 2d_i' \frac{1}{T^2} \sum_{t=2}^{T} e_{it} \widehat{F}_t \\ &= 2\lambda_i' H^{-1} \frac{1}{T^2} \sum_{t=2}^{T} e_{it} V_t + 2d_i' \left(\frac{1}{T^2} \sum_{t=2}^{T} e_{it} (\widehat{F}_t - HF_t) \right) + 2d_i' H \left(\frac{1}{T^2} \sum_{t=2}^{T} e_{it} F_t \right) \\ &= O_p\left(\frac{1}{C_{NT}}\right) + O_p\left(\frac{1}{R_{NT}}\right) O_p\left(\frac{1}{C_{NT}}\right) + O_p\left(\frac{1}{R_{NT}}\right) O_p(1), \end{aligned}$$

where the order of the first term on the right-hand side follows from first using the Cauchy-Schwarz inequality and then (A9), as seen by writing

$$\frac{1}{T^2} \sum_{t=2}^{T} e_{it} V_t \;\leq\; \left(\frac{1}{T^2} \sum_{t=2}^{T} e_{it}^2 \right)^{1/2} \left(\frac{1}{T^2} \sum_{t=2}^{T} ||V_t||^2 \right)^{1/2} \;=\; O_p(1) O_p\left(\frac{1}{C_{NT}}\right).$$

The order of the second term follows by the same argument, after rewriting $\widehat{F}_t - HF_t = -HF_1 + V_t$. The third term is obvious. If follows that III is $O_p(1/C_{NT})$.

Finally, consider part IV. The first term within the parenthesis is $O_p(1/\sqrt{T})$ and can be considered as $O_p(1/R_{NT})$. For the second term, we have

$$\begin{aligned} \frac{1}{T^2} \sum_{t=2}^{T} A_{it} &= \lambda_i' H^{-1} \left(\frac{1}{T^2} \sum_{t=2}^{T} V_t \right) + d_i' H \left(\frac{1}{T^2} \sum_{t=2}^{T} F_t \right) + d_i' \left(\frac{1}{T^2} \sum_{t=2}^{T} (\widehat{F}_t - HF_t) \right) \\ &= O_p\left(\frac{1}{\sqrt{T} C_{NT}}\right) + O_p\left(\frac{1}{R_{NT}}\right) O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{R_{NT}}\right) O_p\left(\frac{1}{\sqrt{T} C_{NT}}\right), \end{aligned}$$

where we make use of (A8) to obtain the individual orders. Thus, by collecting all the terms, we can show that (A7) reduces to

$$\frac{1}{T^2} \sum_{t=2}^{T} \widehat{e}_{it}^2 \;=\; \frac{1}{T^2} \sum_{t=2}^{T} e_{it}^2 - \frac{2}{T^2} \sum_{t=2}^{T} e_{it} A_{it} + O_p\left(\frac{1}{R_{NT}}\right). \tag{A10}$$

By using (A6), (A10), part (b) and the fact that $T(\widehat{\rho}_i - 1) = O_p(1)$, (A1) reduces to

$$\widehat{\sigma}_{\epsilon i}^2 \;=\; \frac{1}{T} \sum_{t=2}^{T} (\Delta e_{it})^2 + O_p\left(\frac{1}{C_{NT}^2}\right) \;\to_p\; \sigma_{\epsilon i}^2. \tag{A11}$$

This result, together with (A10), implies part (a).

Next, consider (b). Again, by definition

$$\widehat{U}_i \;=\; \frac{1}{T} \sum_{t=2}^{T} \widehat{e}_{it-1} \Delta \widehat{e}_{it}.$$

Note that $\widehat{e}_{it}^2 = (\widehat{e}_{it-1} + \Delta\widehat{e}_{it})^2 = \widehat{e}_{it-1}^2 + (\Delta\widehat{e}_{it})^2 + 2\widehat{e}_{it-1}\Delta\widehat{e}_{it}$, from which it follows that

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\widehat{e}_{it-1}\Delta\widehat{e}_{it} &= \frac{1}{2T}\sum_{t=2}^{T}(\widehat{e}_{it}^2 - \widehat{e}_{it-1}^2 - (\Delta\widehat{e}_{it})^2) \\ &= \frac{1}{2T}\widehat{e}_{iT}^2 - \frac{1}{2T}\widehat{e}_{i1}^2 - \frac{1}{2T}\sum_{t=2}^{T}(\Delta\widehat{e}_{it})^2 \qquad \text{(A12)}\end{aligned}$$

and by applying the same trick to e_{it}^2, we have

$$\frac{1}{T}\sum_{t=2}^{T}e_{it-1}\Delta e_{it} = \frac{1}{2T}e_{iT}^2 - \frac{1}{2T}e_{i1}^2 - \frac{1}{2T}\sum_{t=2}^{T}(\Delta e_{it})^2. \qquad \text{(A13)}$$

Now, the terms in the middle of the right-hand side of (A12) and (A13) are clearly $O_p(1/T)$ as $\widehat{e}_{i1} = 0$ while $e_{i1} = O_p(1)$ by assumption. Also, by using (A6) the difference between the third terms is $O_p(1/C_{NT}^2)$. As for the first term, note that

$$\begin{aligned}\frac{1}{T}\widehat{e}_{iT}^2 &= \frac{1}{T}(e_{iT} - e_{i1} - A_{iT})^2 \\ &= \frac{1}{T}e_{iT}^2 + \frac{1}{T}e_{i1}^2 + \frac{1}{T}A_{iT}^2 - \frac{2}{T}e_{iT}A_{iT} - \frac{2}{T}e_{i1}(e_{iT} - A_{iT}) \\ &= \frac{1}{T}e_{iT}^2 + I + II - III - IV, \quad \text{say.} \qquad \text{(A14)}\end{aligned}$$

By using $(a+b)^2 \leq 2(a^2+b^2)$, the triangle inequality, $||AB|| \leq ||A||||B||$ and then (A5), part II can be written as

$$\begin{aligned}|II| &= \frac{1}{T}(\lambda_i' H^{-1}V_T + d_i'\widehat{F}_T)^2 \leq 2||\lambda_i' H^{-1}||^2\left(\frac{1}{T}\sum_{t=2}^{T}||v_t||^2\right) + 2||d_i||^2\left(\frac{1}{T}\sum_{t=2}^{T}||\widehat{f}_t||^2\right) \\ &= O_p\left(\frac{1}{C_{NT}^2}\right) + O_p\left(\frac{1}{R_{NT}^2}\right)O_p(1).\end{aligned}$$

Hence, II is $O_p(1/C_{NT}^2)$.

Part III is simply

$$\begin{aligned}III &= \frac{2}{T}e_{iT}(\lambda_i' H^{-1}V_T + d_i\widehat{F}_T) \\ &= 2\left(\frac{1}{\sqrt{T}}e_{iT}\right)\lambda_i' H^{-1}\left(\frac{1}{\sqrt{T}}V_T\right) + 2\left(\frac{1}{\sqrt{T}}e_{iT}\right)d_i'\left(\frac{1}{\sqrt{T}}\widehat{F}_T\right) \\ &= O_p(1)O_p\left(\frac{1}{C_{NT}}\right) + O_p(1)O_p\left(\frac{1}{R_{NT}}\right)O_p(1),\end{aligned}$$

where the first term on the right-hand side is a direct consequence of (A8) while the second follows from Lemma B.2 (i) of Bai and Ng (2004). Thus, III is $O_p(1/C_{NT})$. Part IV is dominated by $\frac{2}{T}e_{i1}e_{iT}$, which is $O_p(1/\sqrt{T})$ or $O_p(1/R_{NT})$. Therefore, by adding the terms, (A14) simplifies to

$$\frac{1}{T}\hat{e}_{iT}^2 = \frac{1}{T}e_{iT}^2 - \frac{2}{T}e_{iT}A_{iT} + O_p\left(\frac{1}{R_{NT}}\right),$$

which it turn implies (b) and thus the proof of Lemma A.1 is complete. ■

Proof of Theorem 1.

Consider (a). By Lemma A.1 (a) and a first order Taylor expansion of the inverse square root, we get

$$\frac{1}{\sqrt{\hat{V}_i}} = \frac{1}{\sqrt{V_i}} + \frac{1}{V_i^{3/2}}\mathcal{R}_{2i} + O_p\left(\frac{1}{R_{NT}}\right),$$

where

$$\mathcal{R}_{2i} = \sigma_{\epsilon i}^2\frac{1}{T^2}\sum_{t=2}^{T}e_{it-1}A_{it-1}.$$

Let $\mathcal{R}_{1i} = \frac{1}{T}e_{iT}A_{iT}$. Application of Lemma A.1 (b) now gives

$$DF_{\hat{e}}^c(i) = DF_e^c(i) - \frac{1}{\sqrt{V_i}}\mathcal{R}_{1i} + \frac{U_i}{V_i^{3/2}}\mathcal{R}_{2i} + O_p\left(\frac{1}{R_{NT}}\right) = DF_e^c(i) + \mathcal{R}_i,$$

where

$$\mathcal{R}_i = \frac{1}{\sqrt{V_i}}\mathcal{R}_{1i} - \frac{U_i}{V_i^{3/2}}\mathcal{R}_{2i} + O_p\left(\frac{1}{R_{NT}}\right). \tag{A15}$$

Thus, since the first two terms on the right-hand side of (A15) are $O_p(1/C_{NT})$ by Lemma A.1, we have that

$$\begin{aligned}\overline{DF}_{\hat{e}}^c(N) &= \frac{1}{N}\sum_{i=1}^{N}DF_{\hat{e}}^c(i) = \frac{1}{N}\sum_{i=1}^{N}(DF_e^c(i) + \mathcal{R}_i) = \overline{DF}_e^c(N) + \frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_i \\ &= \overline{DF}_e^c(N) + O_p\left(\frac{1}{C_{NT}}\right). \end{aligned} \tag{A16}$$

Consider $\overline{DF}_e^c(N)$. Because $DF_e^c(i) \Rightarrow \mathcal{B}_i$ as $T \to \infty$ and $\mathcal{B}_i$ is independent across i, we obtain the following sequential limit as $T \to \infty$ and then $N \to \infty$

$$\overline{DF}_e^c(N) = \frac{1}{N}\sum_{i=1}^{N}DF_e^c(i) \to_p E(\mathcal{B}), \tag{A17}$$

where the index i in $\mathcal{B}_i$ is suppressed here because of the independence. Now, according to Corollary 1 of Phillips and Moon (1999), since the scaling of $DF_e^c(i)$ is just unity, if we can show that $|DF_e^c(i)|$ is uniformly integrable in T, then (A17) is not only a sequential but also a joint limit as $N, T \to \infty$. But since $DF_e^c(i)$ converges to $\mathcal{B}_i$, we have from Theorem 5.4 of Billingsley (1968) that uniform integrability of $|DF_e^c(i)|$ is equivalent to requiring that

$$E(|DF_e^c(i)|) \;\to\; E(|\mathcal{B}|),$$

which holds trivially since $DF_e^c(i)$ is a scalar so (A17) is indeed a joint limit as $N, T \to \infty$, see Appendix C of Phillips and Moon (1999). This result, together with (A17), imply that

$$\overline{DF}_{\hat{e}}^c(N) \;=\; \overline{DF}_e^c(N) + O_p\left(\frac{1}{C_{NT}}\right) \;\to_p\; E(\mathcal{B}),$$

which establishes (a).

Next, consider (b). We have

$$\begin{aligned}\sqrt{N}(\overline{DF}_{\hat{e}}^c(N) - E(\mathcal{B})) &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N DF_{\hat{e}}^c(i) - E(\mathcal{B})\right) \\ &= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N (DF_e^c(i) + \mathcal{R}_i) - E(\mathcal{B})\right) \\ &= \sqrt{N}(\overline{DF}_e^c(N) - E(\mathcal{B})) + \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathcal{R}_i, \qquad \text{(A18)}\end{aligned}$$

where the last term on the right-hand side can be written as

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathcal{R}_i &= -\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{1}{\sqrt{V_i}}\mathcal{R}_{1i} + \frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{U_i}{V_i^{3/2}}\mathcal{R}_{2i} + O_p\left(\frac{\sqrt{N}}{R_{NT}}\right) \\ &= -I + II + O_p\left(\frac{\sqrt{N}}{R_{NT}}\right), \quad \text{say.} \qquad \text{(A19)}\end{aligned}$$

Note that the reminder vanishes under the condition that $\frac{N}{T} \to 0$ as $N, T \to \infty$.

Consider I. From the proof of Theorem 3 in Bai (2003), we get

$$\begin{aligned}\frac{\sqrt{N}}{\sqrt{T}}A_{it} &= \lambda_i'\left(\frac{1}{N}\sum_{j=1}^N \lambda_j\lambda_j'\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^N \lambda_j \frac{1}{\sqrt{T}}\sum_{s=2}^t \epsilon_{js} \\ &+ \frac{\sqrt{N}}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{s=2}^t f_s'\left(\frac{1}{T}\sum_{s=2}^T f_s f_s'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{s=2}^T f_s\epsilon_{is} + O_p\left(\frac{1}{C_{NT}^2}\right) \\ &= \lambda_i'\left(\frac{1}{N}\sum_{j=1}^N \lambda_j\lambda_j'\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^N \lambda_j \frac{1}{\sqrt{T}}\sum_{s=2}^t \epsilon_{js} + O_p\left(\frac{\sqrt{N}}{\sqrt{T}}\right) + O_p\left(\frac{1}{C_{NT}^2}\right).\end{aligned}$$

Similar to (A.30) in Phillips *et al.* (2001), it is possible to show that as $N \to \infty$

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \lambda_j \sigma_{\epsilon j} W_{\epsilon j}(s) \;\Rightarrow\; \Gamma^{1/2} W_{\epsilon}(s),$$

where $W_{\epsilon j}(s)$ and $W_{\epsilon}(s)$ are independent standard Brownian motions. In view of this result, it is not difficult to see that as $N, T \to \infty$ with $\frac{N}{T} \to 0$

$$\frac{\sqrt{N}}{\sqrt{T}} A_{it} \;\Rightarrow\; \sqrt{\lambda_i' \left(\Sigma^{-1}\Gamma\Sigma^{-1}\right) \lambda_i}\, W_{\epsilon}(s).$$

Let $\phi_i^2 = \lambda_i' \left(\Sigma^{-1}\Gamma\Sigma^{-1}\right) \lambda_i / \sigma_{\epsilon i}^2$. Since $\frac{1}{\sqrt{T}} e_{it} \Rightarrow \sigma_{\epsilon i} W_{\epsilon i}(s)$ as $T \to \infty$, we get

$$\sqrt{N}\mathcal{R}_{1i} \;=\; \sqrt{N}\left(\frac{1}{T} e_{iT} A_{iT}\right) \;=\; \left(\frac{1}{\sqrt{T}} e_{iT}\right)\left(\frac{\sqrt{N}}{\sqrt{T}} A_{iT}\right) \;\Rightarrow\; \sigma_{\epsilon i}^2 \phi_i W_{\epsilon}(1) W_{\epsilon i}(1).$$

Moreover, since $V_i \Rightarrow \sigma_{\epsilon i}^4 \int_0^1 W_{\epsilon i}(s)^2 ds$ as $T \to \infty$, passing $N, T \to \infty$ with $\frac{N}{T} \to 0$

$$\sqrt{N}\left(\frac{1}{\sqrt{V_i}} \mathcal{R}_{1i}\right) \;\Rightarrow\; \phi_i \frac{W_{\epsilon}(1) W_{\epsilon i}(1)}{\sqrt{\int_0^1 W_{\epsilon i}(s)^2 ds}}$$

from which it follows that

$$I \;=\; \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sqrt{V_i}} \mathcal{R}_{1i} \;\Rightarrow\; W_{\epsilon}(1) \left\{ \frac{1}{N} \sum_{i=1}^{N} \phi_i \frac{W_{\epsilon i}(1)}{\sqrt{\int_0^1 W_{\epsilon i}(s)^2 ds}} \right\}.$$

Thus, by assuming that $\frac{1}{N}\sum_{i=1}^{N} \phi_i \to \phi$ as $N \to \infty$, and then applying the same arguments as in equation (A.10) of Phillips *et al.* (2001), we get

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} \phi_i \frac{W_{\epsilon i}(1)}{\sqrt{\int_0^1 W_{\epsilon i}(s)^2 ds}} \;&\to_p\; \phi E\left\{ \frac{W_{\epsilon}(1)}{\sqrt{\int_0^1 W_{\epsilon}(s)^2 ds}} \right\} \\ &= \; \phi E\big[W_{\epsilon}(1)\big] E\left\{ \frac{1}{\sqrt{\int_0^1 W_{\epsilon}(s)^2 ds}} \right\}, \end{aligned} \tag{A20}$$

where the index i is again suppressed because of the independence, and where the equality follows from the fact that $W_{\epsilon}(1)$ is uncorrelated with $\int_0^1 W_{\epsilon}(s)^2 ds$. Note that $E\big[W_{\epsilon}(1)\big]$ is zero, which means that the whole expression is zero. It follows that

$$I \;=\; \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{\sqrt{V_i}} \mathcal{R}_{1i} \;=\; o_p(1). \tag{A21}$$

Next, consider II. By the same arguments used above, as $N, T \to \infty$ with $\frac{N}{T} \to 0$

$$\sqrt{N}\left(\frac{1}{T^2} \sum_{t=2}^{T} e_{it-1} A_{it-1}\right) \;=\; \frac{1}{T} \sum_{t=2}^{T} \left(\frac{1}{\sqrt{T}} e_{it-1}\right)\left(\frac{\sqrt{N}}{\sqrt{T}} A_{it-1}\right) \;\Rightarrow\; \sigma_{\epsilon i}^2 \phi_i \int_0^1 W_{\epsilon i}(u) W_{\epsilon}(u) du.$$

Therefore, because $U_i \Rightarrow \sigma_{\epsilon i}^2 \int_0^1 W_{\epsilon i}(s) dW_{\epsilon i}(s)$ as $T \to \infty$, it is possible to show that

$$\sqrt{N}\left(\frac{U_i}{V_i^{3/2}}\mathcal{R}_{2i}\right) = \sqrt{N}\frac{U_i}{V_i^{3/2}}\left(\sigma_{\epsilon i}^2 \frac{1}{T^2}\sum_{t=2}^{T} e_{it-1}A_{it-1}\right) \Rightarrow \phi_i X_i \left(\int_0^1 W_{\epsilon i}(u) W_\epsilon(u) du\right),$$

where

$$X_i = \frac{\int_0^1 W_{\epsilon i}(s) dW_{\epsilon i}(s)}{\left(\int_0^1 W_{\epsilon i}(s)^2 ds\right)^{3/2}}.$$

This implies that

$$\begin{aligned} II &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{U_i}{V_i^{3/2}}\mathcal{R}_{2i} \Rightarrow \int_0^1 W_\epsilon(r)\left\{\frac{1}{N}\sum_{i=1}^{N}\phi_i X_i W_{\epsilon i}(u)\right\} du \\ &\to_p \int_0^1 W_\epsilon(u)\phi E\big[X W_\epsilon(u)\big] du. \end{aligned} \tag{A22}$$

To find $E\big[X W_\epsilon(r)\big]$, we first derive the joint moment generating function of the triplet

$$(U,\, V,\, S) = \left(\int_0^1 W_\epsilon(s) dW_\epsilon(s),\, \int_0^1 W_\epsilon(s)^2 ds,\, W_\epsilon(u)\right)$$

and then apply an extended version of Lemma 2.3 in Gonzalo and Pitarakis (1998). Towards this end, note that $(U_T,\, V_T,\, S_T) \Rightarrow (U,\, V,\, S)$ as $T \to \infty$ where

$$(U_T,\, V_T,\, S_T) = \left(\frac{1}{T}\sum_{t=1}^{T} x_{t-1}\Delta x_t,\, \frac{1}{T^2}\sum_{t=1}^{T} x_{t-1}^2,\, \frac{1}{\sqrt{T}} x_t\right),$$

where x_t is a cumulated sum of independent standard normals. Now, if we let x denote the T dimensional vector of stacked observations on x_t and $q = \frac{s}{\sqrt{T}} h$, where h is a vector with the value one in the position equal to the integer part of uT and zero otherwise, then the moment generating function of $(U_T,\, V_T,\, S_T)$ is given by

$$\begin{aligned} \varphi_T(u, v, s) &= E\big[\exp(uU_T + vV_T + sS_T)\big] \\ &= \int (2\pi)^{-T/2} \exp\left(uU_T + vV_T + sS_T - \frac{1}{2}\sum_{t=1}^{T}(\Delta x_t)^2\right) dx \\ &= \int (2\pi)^{-T/2} \exp\left(-\frac{1}{2}x' P x + q' x\right) dx = \frac{1}{\sqrt{\det(P)}}\exp\left(\frac{1}{2} q' P^{-1} q\right), \end{aligned}$$

where $P = P_0 + \frac{u}{T} G$ with

$$P_0 = \begin{bmatrix} 2\left(1-\frac{v}{T^2}\right) & -1 & 0 & \cdots & 0 \\ -1 & 2\left(1-\frac{v}{T^2}\right) & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & & 2\left(1-\frac{v}{T^2}\right) & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix},$$

$$G = \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & & 0 \\ \vdots & \ddots & & 2 & -1 \\ 0 & \cdots & 0 & -1 & 0 \end{bmatrix}.$$

Now, consider the identity

$$\frac{1}{V_T^{3/2}} = \frac{1}{\Gamma(3/2)} \left(\int_0^\infty \sqrt{v} \exp(-vV_T) dv \right).$$

This can be used to obtain

$$\begin{aligned} E\left(\frac{U_T S_T}{V_T^{3/2}}\right) &= \frac{1}{\Gamma(3/2)} \int_0^\infty \sqrt{v} \left(\frac{\partial^2}{\partial u \partial s} E\big[\exp(uU_T - vV_T + sS_T)\big] \right) dv \\ &= \frac{1}{\Gamma(3/2)} \int_0^\infty \sqrt{v} \left(\frac{\partial^2}{\partial u \partial s} \varphi_T(u, -v, s) \right) dv, \qquad \text{(A23)} \end{aligned}$$

where the derivatives are taken at $u = s = 0$. To find the required derivative of $\varphi_T(u, -v, s)$, we follow Larsson (1997) and use the following Taylor expansion

$$\varphi_T(u, -v, s) = \frac{1}{\sqrt{\det(P)}} \left(1 - \frac{u}{2T} tr\left(P_0^{-1} G\right) + O\left(\frac{1}{T^2}\right) \right) \left(1 + O\left(\frac{1}{T}\right) \right)$$

from which we deduce that $\frac{\partial^2}{\partial u \partial s} \varphi_T(u, -v, s)$, and hence also (A23), is zero. Thus, letting $T \to \infty$

$$E\left(\frac{U_T S_T}{V_T^{3/2}}\right) \to E\left(\frac{US}{V^{3/2}}\right) = E\big[XW_\epsilon(u)\big] = 0$$

and hence we have shown that

$$II = \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{U_i}{V_i^{3/2}} \mathcal{R}_{2i} = o_p(1). \qquad \text{(A24)}$$

By using (A21) and (A24), it is clear that (A19) reduces to

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathcal{R}_i = -I + II + O_p\left(\frac{\sqrt{N}}{R_{NT}}\right) = o_p(1),$$

which in turn can be inserted into (A17) to obtain

$$\sqrt{N}(\overline{DF}^c_{\hat{e}}(N) - E(\mathcal{B})) \;=\; \sqrt{N}(\overline{DF}^c_e(N) - E(\mathcal{B})) + o_p(1). \tag{A25}$$

Consider the first term on the right-hand side. It holds that

$$\sqrt{N}(\overline{DF}^c_e(N) - E(\mathcal{B})) \;=\; \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(DF^c_e(i) - E(\mathcal{B})),$$

which suggests that in order to show $\sqrt{N}(\overline{DF}^c_e(N) - E(\mathcal{B})) \Rightarrow N(0, \mathrm{var}(\mathcal{B}))$ as N, $T \to \infty$, it is enough to verify that $(DF^c_e(i) - E(\mathcal{B}))$ satisfy conditions (i) to (iv) in Theorem 3 of Phillips and Moon (1999). Conditions (i), (ii) are (iv) are obviously satisfied in view of the fact that the scaling of $DF^c_e(i)$ is again unity. Thus, for this theorem to apply we only need to verify (iii), which requires that $|DF^c_e(i) - E(\mathcal{B})|^2$ is uniformly integrable in T. Towards this end, note that by the continuous mapping theorem, $|DF^c_e(i) - E(\mathcal{B})|^2 \Rightarrow |\mathcal{B}_i - E(\mathcal{B})|^2$ as $T \to \infty$, which together with

$$\begin{aligned} E\left(|DF^c_e(i) - E(\mathcal{B})|^2\right) \;&=\; E\Big((DF^c_e(i) - E(\mathcal{B}))^2\Big) \;\to\; E\Big((\mathcal{B}_i - E(\mathcal{B}))^2\Big) \\ &=\; E\left(|(\mathcal{B}_i - E(\mathcal{B}))|^2\right) \end{aligned}$$

shows that $|DF^c_e(i) - E(\mathcal{B})|^2$ is uniformly integrable in T. Thus, taking the limit as N, $T \to \infty$ with $\frac{N}{T} \to 0$, (A25) becomes

$$\sqrt{N}(\overline{DF}^c_{\hat{e}}(N) - E(\mathcal{B})) \;=\; \sqrt{N}(\overline{DF}^c_e(N) - E(\mathcal{B})) + o_p(1) \;\Rightarrow\; N(0, \mathrm{var}(\mathcal{B})).$$

This completes the proof of (b). ■

Table 1: Simulation results.

			Test size			Remainder	
b	N	T	$\overline{DF}_{\widehat{e}}(N)$	$\overline{DF}_{e}(N)$	$DF_{\widehat{e}}(i)$	Mean	Variance
0	10	100	6.5	5.1	4.8	0.04	40.7
		500	6.6	4.8	4.9	0.02	37.0
		1000	6.6	4.6	4.9	0.01	36.8
	50	100	3.8	4.7	4.8	0.13	14.1
		500	4.4	4.4	4.9	0.02	11.5
		1000	4.8	4.6	4.9	0.01	10.2
5	10	100	19.9	4.9	8.7	−0.59	63.2
		500	19.4	4.7	8.7	−0.60	60.2
		1000	21.1	4.8	8.8	−0.62	58.9
	50	100	6.1	4.4	5.0	0.03	40.4
		500	7.4	4.5	5.0	−0.04	36.7
		1000	7.3	4.8	5.0	−0.06	36.5
10	10	100	29.1	5.0	11.6	−0.96	67.0
		500	32.5	4.4	12.3	−1.12	66.1
		1000	33.4	4.6	12.1	−1.13	64.3
	50	100	21.9	4.4	7.3	−0.77	48.6
		500	15.6	4.7	6.3	−0.51	38.5
		1000	14.4	4.5	6.1	−0.47	36.7
20	10	100	22.8	5.0	9.6	−0.67	62.5
		500	26.0	4.8	10.0	−0.85	62.1
		1000	26.8	4.4	10.1	−0.90	64.7
	50	100	35.8	4.4	9.1	−1.34	49.3
		500	45.2	4.8	9.5	−1.61	49.7
		1000	45.9	4.9	9.5	−1.67	51.2

Notes: The value b refers to the heteroskedasticity of ϵ_{it} with the value zero representing the homoskedastic unit variance case. The leftmost three columns report the size at the 5% level, while the next column report the mean of of the simulated remainder. The rightmost column report the percentage of the total variance in $\overline{DF}_{\widehat{e}}(N)$ that is due to variance in the remainder.

References

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica* **71**, pp. 135-171.

Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica* **72**, pp. 1127-1177.

Breitung, J., and M. H. Pesaran (2005). Unit Roots and Cointegration in Panels. Forthcoming in *The Econometrics of Panel Data.* Mátyás, L., and P. Sevestre (Eds.), Kluwer Academic Publishers.

Dickey, D. A., and W. A. Fuller (1979). Distribution of the Estimator for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* **74**, pp. 427-431.

Gonzalo, J., and J.-Y. Pitarakis (1998). On the Exact Moments of Asymptotic Distributions in an Unstable Ar(1) with Dependent Errors. *International Economic Review* **39**, pp. 71-88.

Kapetanios, G. (2007). Dynamic Factor Extraction of Cross-Sectional Dependence in Panel Unit Root Tests. *Journal of Applied Econometrics* **22**, pp. 313-338.

Larsson, R. (1997). On the Asymptotic Expectations of Some Unit Root Tests in a First Order Autoregressive Process in the Presence of Trend. *Annals of the Institute of Statistical Mathematics* **49**, pp. 585-599.

Phillips, P. C. B., and H. R. Moon (1999). Linear Regression Limit Theory of Nonstationary Panel Data. *Econometrica* **67**, pp. 1057-1111.

Phillips, P. C. B., H. R. Moon and Z. Xiao (2001). How to Estimate Autoregressive Roots Near Unity. *Econometric Theory* **17**, pp. 29-69.