Heteroscedasticity Robust Panel Unit Root Tests

Joakim WESTERLUND

Faculty of Business and Law, School of Accounting, Economics and Finance, Deakin University, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia (j.westerlund@deakin.edu.au)

This article proposes new unit root tests for panels where the errors may be not only serial and/or cross-correlated, but also unconditionally heteroscedastic. Despite their generality, the test statistics are shown to be very simple to implement, requiring only minimal corrections and still the limiting distributions under the null hypothesis are completely free from nuisance parameters. Monte Carlo evidence is also provided to suggest that the new tests perform well in small samples, also when compared to some of the existing tests. Supplementary materials for this article are available online.

KEY WORDS: Common factors; Cross-section dependence; GARCH; Panel data; Unconditional heteroscedasticity.

1. INTRODUCTION

Researchers are by now well aware of the potential hazards involved when using augmented Dickey–Fuller (ADF)-type unit root tests in the presence of unattended structural breaks in the deterministic trend. But breaks in the trend are not the only way in which structural instability may arise. Take, for example, the financial literature concerned with the testing of the efficient market hypothesis, requiring that stock returns are stationary, in which case violations of the otherwise so common homoscedasticity assumption is more of a rule rather than the exception. In fact, as Poon and Granger (2003, p. 481) stated in their review of this literature.

There are several salient features about financial time series and financial market volatility that are now well documented. These include fat tail distributions of risky asset returns, volatility clustering, asymmetry and mean reversion, and comovements of volatilities across assets and financial markets. More recent research finds correlation among volatility is stronger than that among returns and both tend to increase during bear markets and financial crises.

Time-varying volatility is therefore an important source of structural instability. However, as the citation makes clear there is not only the time variation in the volatility of individual stock returns, but also a great deal of similarity in the volatility of different assets and markets, a finding that has been confirmed by numerous studies (see, e.g., McMillan and Ruiz 2009). When it comes to unit root testing, while the latter is typically ignored, the most common way to accommodate the former is to assume that the time variation is in the conditionally variance only (see Seo 1999), which is of course not very realistic. Indeed, Pagan and Schwert (1990), Lorentan and Phillips (1994), Watson (1999), and Busetti and Taylor (2003), to mention a few, all provided strong evidence against the unconditional homoscedasticity assumption for most financial variables, including exchange rates, interest rates, and stock returns.

Of course, unconditional heteroscedasticity is not only a common feature of financial data, but can be found also in most macroeconomic variables. For example, Kim and Nelson (1999), McConnell and Perez Quiros (2000), and Koop and Potter (2000) all showed how the volatility of U.S. gross domestic product (GDP) has declined since the early 1980s. Busetti and Taylor (2003) considered 16 series of industrial production and 17 series of inflation. Consistent with the evidence for GDP, most series are found to exhibit a significant decline in variability around the early 1980s. Similar results were reported by Cavaliere and Taylor (2009a). Chauvet and Potter (2001) reported decreased volatility also in consumption, income, and in aggregate employment, the latter finding is confirmed by results of Warnock and Warnock (2000). Sensier and van Dijk (2004) considered no less than 214 monthly macroeconomic time series. According to their results, around 80% of the series were subject to a significant break in volatility. The evidence also suggests that there is a great deal of similarity in the volatility within groups of series, such as industries, sectors, and real and nominal variables. Hence, again, the unconditional variance is not only time varying, but also co-moving.

The current article can be seen as a response to the above observations. The purpose is to devise a panel-based test procedure that is flexible enough to accommodate not only the “usual suspects” of serial and cross-correlation, but also general (unconditional) heteroscedasticity. The tests should also be simple. In particular, they should not suffer from the same computational drawback as many of the existing time series tests, which typically involve some kind of resampling and/or nonparametric estimation (see Beare 2008; Cavaliere and Taylor 2009b).

The way we accomplish this is by exploiting the information contained in the similarity of both the variance and level of the cross-sectional units, a possibility that has not received any attention in the previous literature. In fact, except for the recent study of Demetrescu and Hanck (2011), this is the only
panel investigation that we are aware of to address the issue of heteroscedasticity.

Two tests based on the new idea are proposed. One is designed to test for a unit root in a single time series, while the other is designed to test for a unit root in a panel of multiple time series. Both test statistics are very convenient in that no prior knowledge regarding the structure of the heteroscedasticity is needed, and still the implementation is extremely simple, requiring only minimal corrections. The asymptotic ticity is needed, and still the implementation is extremely simple, requiring only minimal corrections. The asymptotic

Consider \( \sigma_{m,t}^2 \). It is assumed that the cross-section can be divided into \( M \) cross-section homoscedastic groups with the \( m \)th group containing the units \( i = N_{m-1} + 1, \ldots, N_m \), where \( N_0 = 1 \) and \( N_{M+1} = N \). That is, the heteroscedasticity across the cross-section is made up of \( M \) discrete jumps between otherwise homoscedastic groups. As a matter of notation, in what follows, if a variable, \( x_{m,i} \) say, depends on both \( m \) and \( i \), since the latter index runs across groups, the dependence on \( m \) will be suppressed, that is, \( x_{m,i} \) will be written as \( x_i \). Note also that while the sequential group structure assumed here, in which units \( i = N_{m-1} + 1, \ldots, N_m \) are arbitrarily assigned to group \( m \), is without loss of generality. In fact, as we demonstrate in Section 3.3, in applications it is actually quite useful to think of the cross-section as being ordered.

**Assumption 2.**

(a) \( N_m - N_{m-1} \geq qN \) and \( N_m = \lfloor \lambda_m N \rfloor \), where \( q > 0 \), \( \lambda_0 < \lambda_1 < \cdots < \lambda_M < \lambda_{M+1} \), \( \lambda_0 = 0 \) and \( \lambda_{M+1} = 1 \);
(b) \( \sigma_{m,t}^2 \rightarrow \infty \) as \( T \rightarrow \infty \), where \( \sigma_{m,t}^2 \) and \( \Sigma(s) \) are bounded and square-integrable with a finite number of points of discontinuity;
(c) \( \sum_{i=1}^{N_m} \sigma_{m,t}^2 \Theta_i^2 / N \rightarrow \Lambda(s) > 0 \) as \( N, T \rightarrow \infty \).

The parameter of interest, \( \rho_i \), is assumed to satisfy Assumption 3.

**Assumption 3.**

\[
\rho_i = \exp \left( \frac{c_i}{N^\kappa T} \right),
\]
where \( \kappa \geq 0 \) and \( c_i \) is iid across \( i \) with \( E(c_i) = \mu_c \) and \( \text{var}(c_i) = \sigma_c^2 < \infty \). \( c_i \) is independent of \( \varepsilon_{ij,t} \) and \( F_t \) for all \( i, j, t, \) and \( s \).

Hence, if \( c_i = 0 \), then \( \rho_i = 1 \), and therefore \( y_{ij,t} \) is unit root nonstationary, whereas if \( c_i < 0 \) (\( c_i > 0 \)), then \( \rho_i \) approaches one from below (above) and therefore \( y_{ij,t} \) is “locally” stationary (explosive) (in the mean). The relevant null hypothesis here is given by \( H_0 : c_1 = \cdots = c_N = 0 \), which equivalent to \( H_0 : \mu_c = \sigma_c^2 = 0 \). Similarly, the alternative hypothesis of \( H_1 : c_i \neq 0 \) for at least some \( i \) can be written equivalently as \( H_1 : \mu_c \neq 0 \) and/or \( \sigma_c^2 > 0 \). Assumption 3 therefore greatly reduces the dimensionality of the testing problem (from \( N \) to only two). The “closeness” of the local alternative to the null is determined by \( \kappa \). The most “distant” alternative is obtained by setting \( \kappa = 0 \), in which case \( \rho_i = \exp(c_i/T) \) does not depend on \( N \). This is the typical time series parameterization, which we will be working under when studying the local asymptotic power of our time series statistic (see Section 3.4). As \( \kappa \) increases, we get closer to the null, which means that deviations \( (c_i \neq 0) \) will be more difficult to detect. The panel statistic that we will consider has nonnegligible local power for \( \kappa = 1/2 \) (see Section 3.2) and is therefore able to pick up smaller deviations than the time series statistic.

**Remarks.**

1. The variance of \( \varepsilon_{ij,t} \) is a key in our approach. The assumption of “group-wise heteroscedasticity” in the cross-section

2. The rest of the article is organized as follows. Section 2 introduces the model, while Section 3 presents the test statistics and their asymptotic distributions, which are evaluated in small samples in Section 4. Section 5 concludes.
not only seems like a reasonable scenario in many applications (see Section 3.3 for a discussion), but is also one of the strengths of the approach, as it allows for consistent estimation of the group-specific variances (see Section 3.1). This is facilitated by Assumption 2(a), which states that the groups increase proportionally with $N$. In this regard, it is quite obvious that while in theory $q$ (the proportion) can be made arbitrarily small, in applications the value of $q$ is not irrelevant, as larger groups are expected to increase the precision of the estimated group-specific variances. In Section 4, we discuss the importance of $q$ in small samples (see also the supplement for this article).

2. While the heteroscedasticity across the cross-section needs to have the groupwise structure, the heteroscedasticity across time is essentially unrestricted. In the derivations, we assume that $\sigma_{m,i}^2$ and $\Sigma_i$ are nonrandom, which allows for a wide class of deterministic variance models, including models with breaks and linear trends. However, provided that $\sigma_{m,i}^2$, $\Sigma_i$, $\epsilon_{i,t}$, and $F_t$ are independent with $E(\sigma_{m,i}^2) > 0$ and $E(\Sigma_i) > 0$, $\sigma_{m,i}^2$, and $\Sigma_i$ can also be random. Conditional heteroscedasticity, including generalized autoregressive conditional heteroscedasticity (GARCH), is permitted too, as heteroscedasticity of this form does not affect the asymptotic results.

3. The assumption in Equation (4) is quite common (see, e.g., Pesaran, Smith, and Yamagata 2009), and ensures that the serially uncorrelated error $\epsilon_{i,t}$ has a strict common factor structure. This is more restrictive than the approximate factor model considered by Bai and Ng (2010), where the idiosyncratic error is allowed to be “mildly” cross-correlated, but is necessary for the proofs. Except for this, however, the above model is actually quite general when it comes to the allowable serial and cross-sectional dependencies. Note in particular how the error in Equation (2) admits to a general dynamic factor structure of potentially infinite order.

4. The assumption that the mean and variance of $c_i$ are equal across $i$ is not necessary and can be relaxed as long as the cross-sectional averages of $E(c_i)$ and $var(c_i)$ have limits such as $\mu_c$ and $\sigma_c^2$, respectively. The assumption that $c_i$ is independent of $\epsilon_{i,t}$ and $F_t$ can also be relaxed somewhat. In particular, note that since under the null of a unit root

$$c_1 = \cdots = c_N = 0$$

correlation between $c_i$ and $\epsilon_{i,t}$ and/or $F_t$ is only an issue under the alternative hypothesis, and then it can be shown that in the typical panel parameterization of $\rho_t$ with $\kappa = 1/2$, the effect of correlation between $c_i$ and $\epsilon_{i,t}$ is negligible, although otherwise this is not necessarily so. Unfortunately, such a correlation greatly complicates both the derivation and interpretation of the results, and we therefore follow the usual practice in the literature and assume that Assumption 3 holds.

3. TEST PROCEDURES

3.1 A Time Series Test

The time series test statistic that we consider, henceforth denoted by $\tau_i$, is very simple and can be seen as a version of the Lagrange multiplier (LM) test statistic for a unit root (see Remark 4). The implementation consists of four steps:

1. Obtain $\Delta \hat{R}_{i,t}$ as the ordinary least-squares (OLS) residual from a regression of $\Delta y_{i,t}$ onto $(\Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p})$.
2. Obtain $\hat{F}_i$ and $\hat{\Phi}_i$ by applying the principal components method to $\Delta \hat{R}_{i,t}$. Compute $\hat{\epsilon}_{i,t} = \Delta \hat{R}_{i,t} - \hat{\Phi}_i \hat{F}_i$.
3. Compute $\hat{w}_{m,t} = 1/\hat{\sigma}_{m,t}$, where $\hat{\sigma}_{m,t} = \sum_{i=0}^{m(N_m-1+1)} \hat{e}_{i,t}^2 / (N_m - N_{m-1})$.
4. Compute $\tau_i = \left( \sum_{t=p_i+2}^{T} \Delta \hat{R}_{m,t} \hat{R}_{w,t-1} \right)^2 / \sum_{t=p_i+2}^{T} \hat{R}_{w,t-1}^2$, where $i = N_{m-1} + 1, \ldots, N_m$, $\Delta \hat{R}_{w,t} = \hat{w}_{m,t} \hat{\epsilon}_{i,t}$ and $\hat{R}_{w,t} = \sum_{i=p_i+2}^{T} \Delta \hat{R}_{w,i}$.

Remarks.

5. $\tau_i$ can be seen as a version of the true LM test statistic (based on known parameters) for testing $H_0 : c_i = 0$, which is given by $(\sum_{i=p_i+2}^{T} w_{m,t}^2 \Delta \hat{R}_{i,t} \hat{R}_{i,t-1})^2 / \sum_{i=p_i+2}^{T} w_{m,t}^2 \hat{R}_{i,t-1}$, where $w_{m,t} = 1/\hat{\sigma}_{m,t}$, $\Delta \hat{R}_{i,t} = \hat{\phi}_i(L) \Delta y_{i,t}$ and $R_{i,t} = \sum_{i=p_i+2}^{T} \Delta R_{i,t}$. Unfortunately, the asymptotic distribution of this test statistic depends on $\sigma_{m}(r)$ and $\Sigma(r)$, which means that critical values can only be obtained for a particular choice of $(\sigma_{m}(r), \Sigma(r))$. The use of accumulated weighted first differences (as when replacing $w_{m,t} \hat{R}_{i,t-1}$ with $\sum_{i=p_i+2}^{T} w_{m,t} \Delta R_{i,t}$) eliminates this dependence.

6. The formula for $\tau_i$ reveals some interesting similarities with results obtained previously in the literature. In particular, note how $\tau_i$ is basically the squared, weighted, and “defactored” equivalent of the conventionalADF test. It can also be regarded as a generalization of the LM test developed by Ahn (1993), and Schmidt and Phillips (1992), who considered the problem of testing for a unit root in a pure time series setting with homoscedastic errors. The way that $\Delta \hat{R}_{w,i}$ is accumulated up to levels is very similar to the approach of Bai and Ng (2004). But while in that article the accumulation is just an artifact of their proposed correction for cross-section dependence, here it is used as a means to recursively weight the observations, and ultimately to remove the effect of the heteroscedasticity (see Beare 2008, for a similar approach).

7. As is well known, under certain conditions (such as normality), the LM test statistic possesses some optimality properties. While the test statistic considered here is not the true LM statistic, it is similar and is therefore expected to inherit some of these properties. Apart from the resemblance with the true LM statistic, the use of a squared test statistic has the advantage that it does not rule out the possibility of explosive units ($c_i > 0$), which seems like a relevant scenario in many applications, especially in financial economics, where data can exhibit explosive behavior. Explosive behavior is
also more likely if N is large, which obviously increases the probability of extreme events regardless of the application considered. This is particularly relevant when constructing pooled panel test statistics (see Section 3.2), in which case positive and negative values of $c_i$ may otherwise cancel out, causing low power.

8. $\tau_i$ can be modified to incorporate information regarding the direction of the alternative. Suppose, for example, that one would like to test $H_0 : c_i = 0$ versus $H_1 : c_i < 0$. A very simple approach would be to first test $H_0$ versus $H_1$. If $H_0$ is accepted, then we conclude that $y_{it}$ unit root nonstationary and proceed no further, whereas if $H_0$ is rejected, then the testing continues by checking the sign of $\sum_{t=1}^{T} \Delta R_{wi,t} \hat{R}_{wi,t-1}$. Only if the sign is negative do we conclude that the evidence is in favor of stationarity and not of explosiveness. Of course, because of the data snooping, this test would not be correctly sized. A more appropriate testing approach involves replacing $\tau_i$ with $1(\sum_{t=1}^{T} \Delta R_{wi,t} \hat{R}_{wi,t-1} / T < 0) \tau_i$, where $1(A)$ is the indicator function for the event A. The asymptotic null distribution of the resulting joint test statistic can be worked out by using the results of Abadir and Distaso (2007).

9. In regression analysis with heteroscedastic errors it is common practice to use $\hat{\sigma}^2_{m,i}$ to estimate $\sigma^2_{m,i}$. However, as Lopez (2001) pointed out, while unbiased, because of its asymmetric distribution, $\hat{\sigma}^2_{i,t}$ is a very imprecise estimator of $\sigma^2_{m,i}$. $\hat{\sigma}^2_{m,i}$ uses more information and is therefore expected to perform much better. Beare (2008) used a similar recursive weighting scheme as the one considered here, but where $\sigma^2_{m,i}$ is estimated in a nonparametric fashion, which not only leads to a very low rate of consistency but also complicates implementation.

The asymptotic distribution of $\tau_i$ is given in the following theorem.

**Theorem 1.** Under the conditions laid out in Section 2, given $\kappa = 0$, as $N, T \to \infty$

$$\tau_i \to_d \left( \frac{\int_{s=0}^{1} V_{R,i}(s)dV_{R,i}(s)}{\int_{s=0}^{1} V_{R,i}(s)^2 ds} \right)^2,$$

where

$$w_m(s) = \frac{1}{\sigma_m(s)},$$

$$V_{R,i}(s) = W_{e,i}(s) + c_i \int_{r=0}^{s} w_m(r) B_{R,i}(r) dr,$$

$$B_{R,i}(s) = \Theta_i'B_{F,i}(s) + B_{e,i}(s),$$

$$B_{e,i}(s) = \int_{r=0}^{s} \exp(s-r)c_i\sigma_m(r)dW_{e,i}(r),$$

$$B_{F,i}(s) = \int_{r=0}^{s} \exp(s-r)c_i\Sigma(r)^{1/2}dW_{F}(r),$$

with $\to_d$ signifying convergence in distribution, $i = N_{m-1} + 1, \ldots, N_m$, and $W_{e,i}(s)$ and $W_{F}(s)$ being two independent standard Brownian motions.

To appreciate fully the implication of Theorem 1, note that since $V_{R,i}(s)$ satisfies the differential equation $dV_{R,i}(s) = dW_{e,i}(s) + c_i w_m(s) B_{R,i}(s) ds$, the numerator of the asymptotic test distribution can be expanded as

$$\int_{s=0}^{1} V_{R,i}(s)dV_{R,i}(s) = \int_{s=0}^{1} V_{R,i}(s)w_m(s)B_{R,i}(s)ds + \int_{s=0}^{1} V_{R,i}(s)dW_{e,i}(s).$$

This illustrates how the presence of $c_i$ has two effects. The first is to shift the mean of the test statistic and is captured by the first term on the right-hand side, whereas the second is to affect the variance and is captured by $V_{R,i}(s)$ (which depends on $c_i$). It also illustrates how the local asymptotic power depends on both the heteroscedasticity and the cross-section dependence, as captured by $\sigma^2_m(s)$ and $\Theta_i'B_{F,i}(s)$, respectively. Whether this dependence implies higher local power in comparison to the case with homoscedastic errors is not possible to determine unless a particular choice of $\sigma^2_m(s)$ and $\Theta_i'B_{F,i}(s)$ is made. However, we see that if $\sigma^2_m(s) = \sigma^2_m$ and $\Theta_i = 0$, then the dependence on these nuisance parameters disappears. We also see that whenever $\sigma^2_m(s)$ is not constant and/or $\Theta_i \neq 0$, then this is no longer the case (if $c_i \neq 0$). Note in particular that even if $\sigma^2_m(s) = \sigma^2_m$, unless $\Theta_i = 0$, the dependence on $\sigma^2_m$ will not disappear. In Section 3.2, we elaborate on the power implications of the heteroscedasticity.

Under the unit root hypothesis that $c_i = 0$, $dV_{R,i}(s) = dW_{e,i}(s)$, and hence $V_{R,i}(s) = W_{e,i}(s)$, which in turn implies

$$\tau_i \to_d \left( \frac{\int_{s=0}^{1} W_{e,i}(s)dW_{e,i}(s)}{\int_{s=0}^{1} W_{e,i}(s)^2 ds} \right)^2$$

as $N, T \to \infty$. Hence, the asymptotic null distribution of $\tau_i$ does not depend on $\sigma_m(s)$ or $\Theta_i'B_{F,i}(s)$. It is therefore completely free of nuisance parameters. In fact, closer inspection reveals that the asymptotic distribution of $\tau_i$ is nothing but the squared ADF test distribution, which has been studied before by, example, Ahn (1993), who also tabulated critical values (Table 1).

| $N$ | $T$ | Case | $\tau$ | $\tau_{IV}$ | $\delta_{NT}$ | $\tau_i$ | $\tau_{IV,i}$ |
|-----|-----|------|--------|-------------|---------------|----------|--------------|
| 10  | 100 | 1    | 7.0    | 0.8         | 3.0           | 5.3      | 1.3          |
| 20  | 400 | 1    | 5.6    | 0.1         | 3.9           | 5.1      | 0.4          |
| 10  | 100 | 2    | 7.7    | 1.4         | 2.7           | 5.6      | 2.2          |
| 20  | 400 | 2    | 5.6    | 0.3         | 3.7           | 5.2      | 0.8          |
| 10  | 100 | 3    | 8.0    | 1.5         | 3.0           | 5.6      | 2.3          |
| 20  | 400 | 3    | 5.6    | 0.3         | 3.8           | 5.2      | 0.8          |

*Notes:* Variance cases 1–3 refer to homogeneity, a discrete break, and a smooth transition break, respectively. $\phi_i$ refers to the first-order autoregressive serial correlation coefficient in the errors.
Remarks.

10. The fact that the asymptotic null distribution of \( \tau_i \) is free of nuisance parameters is a great operational advantage that is not shared by many other tests. Consider, for example, the (adaptive) test statistic analyzed by Boswijk (2005), which allows for heteroscedasticity of the type considered here (but no cross-section dependence). However, unlike \( \tau_i \), the asymptotic distribution of this test statistic depends on \( \sigma^2_m(r) \), which is of course never known in practice. Similarly, while the unit root tests of Pesaran (2007) and Pesaran, Smith, and Yamagata (2009) do allow for cross-section dependence in the form of common factors (but no heteroscedasticity), their asymptotic null distributions depend on \( B_{F,i}(s) \), whose dimension is generally unknown, which greatly complicates the implementation. In fact, the only test statistic that we are aware of that comes close to ours in terms of generality and still having a completely nuisance parameter free null distribution is the Cauchy-based \( t \)-statistic of Demetrescu and Hanck (2011).

11. The new test statistic can be applied under very general conditions when it comes to the dynamics and heteroscedasticity of the errors. However, it cannot handle models with a unit-specific trend that needs to be estimated. The reason is that while the effect of the estimation of the intercept is negligible, this is not the case with the trend, whose nuisance parameters is a great operational advantage that is not shared by many other tests. Consider, for example, the (adaptive) test statistic analyzed by Boswijk (2005), which allows for heteroscedasticity of the type considered here (but no cross-section dependence). However, unlike \( \tau_i \), the asymptotic distribution of this test statistic depends on \( \sigma^2_m(r) \), which is of course never known in practice. Similarly, while the unit root tests of Pesaran (2007) and Pesaran, Smith, and Yamagata (2009) do allow for cross-section dependence in the form of common factors (but no heteroscedasticity), their asymptotic null distributions depend on \( B_{F,i}(s) \), whose dimension is generally unknown, which greatly complicates the implementation. In fact, the only test statistic that we are aware of that comes close to ours in terms of generality and still having a completely nuisance parameter free null distribution is the Cauchy-based \( t \)-statistic of Demetrescu and Hanck (2011).

### 3.2 A Panel Data Test

Given that the asymptotic null distribution of \( \tau_i \) is free of nuisance parameters, the various panel unit root tests developed in the literature for the case of homoscedastic and/or cross-correlation free errors can be applied also to the present more general case. In this section, we consider the following panel version of \( \tau_i \):

\[
\tau = \frac{\left( \sum_{j=1}^{N} \sum_{t=1}^{T} \Delta \hat{R}_{i,j} \right)^2}{\sum_{j=1}^{N} \sum_{t=p_{i+2}}^{T} \hat{R}_{i,j}^2 - 1}.
\]

The main reason for looking at a between rather than a within type test statistic is that such test statistics are known to have relatively good local power properties (see Westerlund and Breitung 2012). The asymptotic distribution of \( \tau \) is given in Theorem 2.

**Theorem 2.** Under the conditions of Theorem 1, given \( \kappa = 1/2 \), as \( N \rightarrow \infty \) with \( N/T \rightarrow 0 \),

\[
\tau \rightarrow d \chi^2(0^2),
\]

where

\[
\theta = \lim_{N \rightarrow \infty} \frac{\sum_{m=1}^{M} (\lambda_m - \lambda_{m-1}) f_{r,0}^{T} \bar{q}_{N,m}(r)dr}{\sqrt{\sum_{m=1}^{M} (\lambda_m - \lambda_{m-1}) f_{r,0}^{T} \bar{p}_{N,m}(r)dr}},
\]

\[
\bar{q}_{N,m}(u) = \frac{1}{N_m - N_{m-1}} \sum_{i=N_{m-1}+1}^{N_m} q_{N,i}(u),
\]

\[
\bar{p}_{N,m}(u) = \frac{1}{N_m - N_{m-1}} \sum_{i=N_{m-1}+1}^{N_m} p_{N,i}(u),
\]

\[
p_{N,i}(r) = r + 2 \mu_c \sqrt{N} \int_{u=0}^{T} g_m(u)du + \left( \frac{\mu^2_c + \sigma^2_m}{N} \right) \times \int_{u=0}^{T} \left( g_i(u, v) + 2 \int_{v=0}^{u} g_i(u, v)dv \right)du,
\]

\[
q_{N,i}(u) = \mu_c g_m(u) + \left( \frac{\mu^2_c + \sigma^2_m}{\sqrt{N}} \right) \int_{u=0}^{T} g_i(u, v)dv,
\]

\[
g_m(v) = w_m(v) \int_{u=0}^{T} \sigma_m(r)dr,
\]

\[
g_i(u, v) = w_m(v) \int_{u=0}^{T} \left( \Theta_i(x) + \sigma_m(x) \right)dx,
\]

and \( \chi^2(d) \) is a noncentral chi-square distribution with \( k \) degrees of freedom and noncentrality parameter \( d \).

In agreement with the results reported in Theorem 1 the asymptotic local power of \( \tau \) depends on both the heteroscedasticity and the cross-section dependence, as captured by \( \sigma^2_m(s), \Sigma(s), \) and \( \Theta_i \). In fact, Theorem 2 is the first result that we are aware of that shows how power is affected by all three parameters.\(^5\)

The first thing to note is that while this is not the case for \( \sigma^2_m(s) \), the effect of \( \Theta_i \) and \( \Sigma(s) \) is negligible (as \( N \rightarrow \infty \)). This means that our defactoring approach is effective in removing the effect of the common component, not only under the null but also under the local alternative. As in Section 3.1, unless a particular choice of \( \sigma^2_m(s) \) is made, its effect on power cannot be determined. However, we see that if \( N < \infty \), unless \( \Theta_i = 0 \), the dependence on \( \sigma^2_m(s) \) will not disappear even if \( \sigma^2_m(s) = \sigma^2_m \), which is again in agreement with Theorem 1.

\(^4\)A detailed proof is available upon request.

\(^5\)Some confirmatory Monte Carlo results are available upon request.
To illustrate how a particular choice of \( \sigma_m^2(s) \) can affect power let us consider as an example the case when \( \sigma_m(s) = \sigma(s) = 1 + 1(s > b)/4 \). With \( b = 1/2 \) this is the discrete break case considered in the simulations (see Section 4). Clearly,

\[
\int_{v=0}^{1} w(v) \int_{v=0}^{v} \sigma(r) dv d\nu = \int_{v=0}^{b} v dv + \int_{v=b}^{1} (4b + (v - b)) dv = \frac{1}{2} + 3b(1 - b),
\]

suggesting that \( \int_{r=0}^{1} \bar{\sigma}_{N,m}(r) dr \rightarrow \mu_c(1/2 + 3b(1 - b)) \) as \( N \rightarrow \infty \). But we also have that \( \int_{r=0}^{1} \bar{\sigma}_{N,m}(r) dr \rightarrow \int_{r=0}^{1} r dr = 1/2 \), from which we obtain \( \theta = \mu_c\sqrt{2(1/2 + 3b(1 - b))} \). Now, the minimal value of \( \theta^2 \) is obtained by setting \( b = 0 \) or \( b = 1 \), in which case \( \sigma(s) = 1 \). It follows that in this particular example \( \theta^2 \), and hence also power, is going to be higher with heteroscedasticity \( (b \in (0, 1)) \) than without \( (b = 0 \text{ or } b = 1) \).

The maximal value of \( \theta^2 \) is obtained by setting \( b = 1/2 \).

Under the null hypothesis that \( \mu_c \) and all its moments are zero, \( \bar{\sigma}_{N,m}(u) = 0 \), suggesting that the result in Theorem 2 reduces to

\[
\tau \rightarrow a \chi^2(0) \sim \chi^2_1,
\]

a central chi-squared distribution with one degree of freedom.

Remarks.

12. While most research on the local power of panel unit root test assume that \( N \rightarrow \infty \) (and also \( T \rightarrow \infty \)) and only report results for the resulting first-order approximate power function (see, e.g., Moon, Perron, and Phillips 2007), which only depends on \( \mu_c \), Theorem 2 also accounts for \( \sigma_m^2 \), and is therefore expected to produce more accurate predictions, a result that is verified using Monte Carlo simulation in Section 4. The theorem also shows how power is driven mainly by \( \mu_c \), and that the effect of \( \sigma_m^2 \) goes to zero as \( N \rightarrow \infty \).

13. While nonnegligible for \( \kappa = 0 \), \( \tau_i \) has negligible local power for \( \kappa > 0 \). Intuitively, when \( \kappa > 0 \), the deviations in \( \rho_i \) from the hypothesized value of one are too small to detect them. The fact that \( r \) has negligible local power for \( \kappa = 1/2 \) means that it is able to pick-up deviations that are undetectable by \( \tau_i \). Thus, as expected, the use of the information contained cross-sectional dimension leads to increased power. The fact that the assumptions are the same (except for the requirement that \( N/T \) should go to zero), suggests that this power advantage does not come at a cost of added (homogeneity) restrictions.

14. The condition that \( N/T \rightarrow 0 \) as \( N \rightarrow \infty \) is standard even when testing for unit roots in homoscedastic and cross-sectionally independent panels. The reason for this is the assumed cross-sectional heterogeneity of the DGP, whose elimination induces an estimation error in \( T \), which is then aggravated when pooling across \( N \). The condition that \( N/T \rightarrow 0 \) as \( N \rightarrow \infty \) prevents this error from having a dominating effect (see Westerlund and Breitung 2011, for a detailed discussion). Demetrescu and Hanck (2011) considered a setup that is similar to ours. However, they require \( N/T^{1/5} \rightarrow 0 \), which obviously puts a limit on the allowable sample sizes. For example, if \( N = 10 \), then \( T > 100,000 \) is required for \( T > N^0 \) to be satisfied. Hence, unless one is using long-span high-frequency data, this approach is not expected to perform well.

3.3 Issues of Implementation

3.3.1 Uncertainty Over the Groups. A problem in applications is how to pick the groups for which to compute the variances. The most natural approach is to exploit if there is a natural grouping of the cross-section. For example, a common case is when we have repeated observations on each of a number of countries, industries, sectors, or regions, where it may be reasonable to presume that the variance is constant within groups but potentially different across.

If there is uncertainty regarding the equality of the groupwise variances, then the appropriate approach will depend on the extent to which there is a priori knowledge regarding the grouping. In this section, we assume that the researcher has little or no such knowledge; in the supplement we consider the case when the researcher knows which units that belong to which group, such that the problem reduces to testing the equality of the groupwise variances. If nothing is known, then the number of groups and their members can be estimated by quasi-maximum likelihood (QML). To fix ideas, suppose that \( M = 1 \), such that there are only \( M = 1 \) groups. The problem of determining the groups therefore reduces to the problem of consistent estimation of the group threshold \( N_i \). For this purpose, it is convenient to treat \( \hat{\delta}^2_{1,i} \) and \( \hat{\delta}^2_{2,i} \) as functions of \( n = qN, \ldots, (1-q)N \), that is, \( \hat{\delta}^2_{1,i}(n) = \sum_{i=1}^{n} \hat{\delta}^2_{1,i}/n \) and \( \hat{\delta}^2_{2,i}(n) = \sum_{i=q+1}^{N} \hat{\delta}^2_{1,i}/(N-n) \). In this notation, the QML objective function is given by

\[
QML(n) = \sum_{t=p+1}^{T} \left[ n \log \left( \hat{\delta}^2_{1,i}(n) + (N-n) \log \left( \hat{\delta}^2_{2,i}(n) \right) \right) \right],
\]

and the proposed estimator of \( N_i \) is given by

\[
\hat{N}_1 = \arg \min_{n=qN, \ldots, (1-q)N} \text{QML}(n).
\]

Proposition 1 shows that \( \hat{N}_1 \) is consistent for \( N_i \).

\[\text{Proposition 1.} \] Under the conditions of Theorem 1, given \( \kappa \geq 0 \), as \( N, T \rightarrow \infty \) with \( N/T \rightarrow 0 \)

\[P(\hat{N}_1 = N_i) \rightarrow 1.\]

The proposed QML estimator of \( \hat{N}_1 \) is analogous to the OLS-based breakpoint estimator considered by, for example, Bai (1997), and Bai and Perron (1998) in the time series case. The reason for using QML is that OLS can only be used in case of a break in the mean. To implement the minimization, the following two-step approach may be used.

1. Compute \( \hat{\delta}^2_{1,i} = \sum_{t=p+1}^{T} \hat{\delta}^2_{1,i}/T \) for each \( i \), and order the cross-section units accordingly.
2. Obtain \( \hat{N}_1 \) by grid search at the minimum QML(n) based on ordered data.

If \( M > 1 \), then the one-at-a-time approach of Bai (1997) may be used. The objective function is therefore identical to QML(n). Let \( \hat{N}_1 \) be the value that minimizes QML(n). \( \hat{N}_1 \) is not necessarily estimating \( N_i \); however, it is consistent for one of the thresholds. Once \( \hat{N}_1 \) has been obtained, the sample is split at
this estimate, resulting in two subsamples. We then estimate a single threshold in each of the subsamples, but only the threshold associated with the largest reduction QML(n) is kept. Denote the resulting estimator as $\hat{N}_2$. If $\hat{N}_2 < \hat{N}_1$, then we switch subscript so that $\hat{N}_1 < \hat{N}_2$. Now, if $M = 2$, the procedure is stopped here. If, on the other hand, $M > 2$, then the third threshold is estimated from the three subsamples separated by $\hat{N}_1$ and $\hat{N}_2$. Again, a single break point is estimated for each of the subsamples, and the one associated with the largest reduction in the objective function is retained. This procedure is repeated until estimates of all $M$ thresholds are obtained. When $M$ is unknown, before an additional threshold is added, we test if this leads to a reduction in the objective function (see Bai and Perron 1998, for a similar approach).\(^7\)

Remarks.

15. In contrast to conventional data clustering, which is computationally very demanding, the above minimization approach is very fast. The key is the use of $\hat{\alpha}_j^2$ as an ordering device for the cross-section (see Lin and Ng 2012, for a similar approach). This makes the grouping problem analogous to the problem of breakpoint estimation in time series, for which there is a large literature (see Perron 2006, for an overview). The one-at-a-time approach of Bai (1997) is computationally very convenient and leads to tests with good small-sample performance (see Section 4). Note also that the ordering can be done while ignoring the heteroscedasticity in $\tau$ (see Lin and Ng 2012).

16. An alternative to the above break estimation approach to the grouping is to use $K$-means clustering (see, e.g., Lin and Ng 2012). In this case, we begin by randomly initializing the grouping. Each unit is then reassigned, one at a time, to the other groups. If the value of the objective function decreases, the new grouping is kept; otherwise, one goes back to the original grouping. This procedure is repeated until no unit changes group. In our simulations, the proposed approach was not only faster but also led to huge gains in performance when compared to clustering.

3.3.2 Lag Selection. It is well known that if the true lag order $p_t$ is unknown but the estimator $\hat{p}_t$ is such that $\hat{p}_t \to \infty$ and $\hat{p}_t = o(T^{1/3})$, then conventional unit root test statistics have asymptotic distributions that are free of nuisance parameters pertaining to the serial correlation of the errors, even if $p_t$ is infinite and the errors are (unconditionally) heteroscedastic (Cavaliere and Taylor 2009b). One possibility is therefore to set $\hat{p}_t$ as a function of $T$. Of course, in practice it is often preferable to use a data-driven approach. In this context, Pötscher (1989) showed that if the errors are unconditionally heteroscedastic, the consistency of the conventional Bayesian information criterion (BIC) cannot be guaranteed. In this subsection, we therefore propose a new information criterion that is robust in this regard. The idea is the same as that of Cavaliere et al. (2012), that is, rather than modifying the information criterion (applied to the original data), we use the heteroscedasticity corrected variables $\hat{R}_{i,t-1}$ and $\Delta \hat{R}_{i,t}$ as input. In particular, the following information criterion, which can be seen as a version of the modified BIC (MBIC) of Ng and Perron (2001), is considered:

$$\text{MBIC}_i(p) = \log(\hat{\sigma}^2(p)) + \frac{\ln(T)}{T}(p + \tau(k)),$$

where $\tau(k) = \hat{\gamma}^2_i(p)\sum_{t=p+2}^{T} \hat{R}_{i,t-1}^2/\hat{\sigma}^2_i(p)$, $\hat{\gamma}^2_i(p)$ is the OLS slope estimator of $\hat{R}_{i,t-1}$ in a regression of $\Delta \hat{R}_{i,t}$ onto $(\hat{R}_{i,t-1}, \Delta \hat{R}_{i,t-1}, \ldots, \Delta \hat{R}_{i,t-p})$ and $\hat{\sigma}^2_i(p)$ is the estimated sample variance from that regression. The maximum $p$ to be considered is denoted as $p_{\max}$ and is assumed to satisfy $p_{\max} \to \infty$ and $p_{\max} = o(T^{1/3})$. The estimator of $p_t$ is given by

$$\hat{p}_t = \arg \min_{p=0,\ldots,p_{\max}} \text{MBIC}_i(p).$$

In Section 4, we use Monte Carlo simulations to evaluate the performance of the tests based on this lag selection procedure in small samples.

3.3.3 Selection of the Number of Common Factors. In the Appendix, we verify that Assumption C of Bai (2003), and Bai and Ng (2002) is satisfied, suggesting that neither the estimation of the common component, nor the selection of the number of common factors is affected by the heteroscedasticity. Therefore, if the number of common factors, $\kappa$, is unknown, as is typically the case in practice, any of the information criteria proposed by Bai and Ng (2002) can be applied to $\Delta \hat{R}_{i,t}$, giving $\hat{\kappa}$.

If $p_t, \kappa$ and the groups are all unknown, then a joint procedure should be used. In the simulations, we begin by estimating $r$ given a maximum lag augmentation of $p_{\max}$. Once $r$ has been estimated, the MBIC is applied to obtain $\hat{p}_t$ for each unit. Given $\hat{r}$ and $\hat{p}_1, \ldots, \hat{p}_N$, the groups are selected by applying QML to $\hat{\kappa}_{i,t}$.

4. SIMULATIONS

In this section, we investigate the small-sample properties of the new test through a small simulation study using (1)–(5) as DGP, where we assume that $a_1 = 1$, $u_{i,t} = 0$, $\phi(L) = 1 - \phi_1 L$, $\Theta_i \sim N(1, 1)$, $F_i \sim N(0, 1)$, $e_{i,t} \sim N(0, 1)$ and $c_i \sim U(a, b)$.

Three cases regarding the time-variation in $\sigma^2_{m,t}$ are considered: (1) $\sigma^2_t = 1$; (2) $\sigma^2_t = 1 - 1/(t \geq [T/2])3/4$; (3) $\sigma^2_t = 1 - 3/4(1 + \exp(-(t - [T/2])))$. While case 2 corresponds to a single discrete break at time $[T/2]$ when $\sigma^2_t$ changes from 1 to 1/4, case 3 corresponds to a smooth transition break from 1 to 1/4 with $[T/2]$ being the transition midpoint. Thus, case 2 is basically the finite-sample analogue of the example given in Section 3.2. Cross-section variation is induced by setting $\sigma^2_{m,t} = \sigma^2_t(1 + (i > [N/2]))$. Hence, in this DGP there are two groups. In the first, containing the first $[N/2]$ units, $\sigma^2_{m,t} = \sigma^2_t$, while in the second, containing the remaining units, $\sigma^2_{m,t} = 2\sigma^2_t$. The data are generated for 5,000 panels with $T = N^2$ (reflecting the theoretical requirement that $N/T$ should go to zero).

For the sake of comparison, the $\tau_{\Delta 1}$ and $\tau_{\Sigma 1}$ test statistics of Demetrescu and Hanck (2011) are also simulated.\(^8\) Both

---

\(^7\)Bai (1997) suggested performing a parameter constancy test in each subsample before the search for additional thresholds is initiated. While such a test can be performed using the test statistics studied in the supplement, unreported Monte Carlo results suggest that the proposed QML-only method leads to best small-sample performance. It is also relatively simple to implement.

\(^8\)Following the recommendation of Demetrescu and Hanck (2011), two versions of $\tau_{\Sigma 1}$ were considered. The first is the original test statistic also considered by
statistics are constructed as \( t \)-ratios of the null hypothesis of a unit root. The difference is that while the first is a time series test, the second is a panel test. Thus, the most relevant comparison here is between \( \tau_i \) and \( \tau_{IV,i} \), on the one hand, and between \( \tau \) and \( \tau_{IV} \), on the other hand. Moreover, in view of the apparent robustness of the sign function to heteroscedasticity (see Demetrescu and Hanck 2011, Proposition 1), the sign-based \( \hat{S}_{NT} \) statistic, which in construction is very close to \( \tau_{IV} \), is also simulated. A number of other tests were attempted (including the ADF\(_2^I(i)\) and \( P^I_2 \) statistics of Bai and Ng 2004) but not included, as their performance was dominated by the performance of the above tests.\(^9\)

In the simulations, the groups and lag augmentation order are selected as described in Section 3.3, using the QML and MBIC, respectively. Consistent with the results of Ng and Perron (1995), the maximum number of lags is set equal to \( p_{\text{max}} = \lfloor 4(T/100)^{2/9} \rfloor \). All tests are based on the same lag augmentation. The trimming parameter in the QML is set to \( q = 0.15 \) (see, e.g., Bai 1997). The number of factors are determined using the \( IC_1 \) information criterion of Bai and Ng (2002) with the maximum number of factors set to \( \lfloor 4(\min\{N, T\})/100^{2/9} \rfloor \) (Bai and Ng 2004).

The results are reported in Tables 1–3. Table 1 contains the size results, while Tables 2 and 3 contain the power results. The information content of these tables may be summarized as follows.

- As expected, \( \tau_i \) and \( \tau \) are generally correctly sized. Of course, the accuracy is not perfect, and some distortions remain. In particular, both tests seem to have a tendency to reject too often. Fortunately, the distortions are never very large, and they vanish quickly as \( N \) and \( T \) increase, which is just as expected, because \( T > N \) (corresponding to the requirement that asymptotically \( N/T \) should go to zero) in the simulations. We also see that none of the tests seem to be affected much by the specification of \( \sigma^2_i \).
- The most serious distortions are found for \( \tau_{IV} \), which is generally quite severely undersized, especially when \( \phi_1 \) is 0.5 and the errors are serially correlated. Of course, with \( T = N^2 \) the requirement that \( T > N^5 \) (corresponding to \( N/T^{1/5} \to 0 \) as \( N \to \infty \)) is not even close to being satisfied, and therefore the poor performance of \( \tau_{IV} \) in this case does not come as a surprise. There is also no tendency for the distortions to become smaller as \( N \) and \( T \) increases. In fact, the distortions actually increase with the sample size. \( \tau_{IV,i} \) is also undersized, although not to the same extent as \( \tau_{IV} \).
- When \( \kappa = 0 \) the power of \( \tau \) is increasing in the sample size, which is to be expected, as the rate of shrinking of the local alternative in this case is not fast enough to prevent the test statistic from diverging with \( N \). The same reasoning

\[ \begin{array}{cccccccc}
 N & T & \text{Case} & \tau & \tau_{IV} & \hat{S}_{NT} & \tau_i & \tau_{IV,i} \\
 \hline
 10 & 100 & 1 & \tau & 31.8 & 3.6 & 8.1 & 6.2 & 2.0 \\
 20 & 400 & 1 & \tau & 64.6 & 1.1 & 14.6 & 6.4 & 0.5 \\
 10 & 100 & 2 & \tau & 45.4 & 5.9 & 7.2 & 6.4 & 3.4 \\
 20 & 400 & 2 & \tau & 88.0 & 2.3 & 13.9 & 6.6 & 1.1 \\
 10 & 100 & 3 & \tau & 45.8 & 6.2 & 7.4 & 6.5 & 3.4 \\
 20 & 400 & 3 & \tau & 88.1 & 2.3 & 14.4 & 6.6 & 1.1 \\
 10 & 100 & 1 & \tau & 66.8 & 9.7 & 15.2 & 8.4 & 2.6 \\
 20 & 400 & 1 & \tau & 98.4 & 5.0 & 28.9 & 9.2 & 0.7 \\
 10 & 100 & 2 & \tau & 73.8 & 11.6 & 12.7 & 8.1 & 4.3 \\
 20 & 400 & 2 & \tau & 99.4 & 7.1 & 23.8 & 8.9 & 1.5 \\
 10 & 100 & 3 & \tau & 74.6 & 12.0 & 12.0 & 8.2 & 4.4 \\
 20 & 400 & 3 & \tau & 99.5 & 7.1 & 24.0 & 8.9 & 1.5 \\
 10 & 100 & 1 & \tau & 61.6 & 9.5 & 15.0 & 8.2 & 2.7 \\
 20 & 400 & 1 & \tau & 95.5 & 5.4 & 28.3 & 9.1 & 0.7 \\
 10 & 100 & 2 & \tau & 66.1 & 11.5 & 11.8 & 7.4 & 4.3 \\
 20 & 400 & 2 & \tau & 96.3 & 7.4 & 23.8 & 8.6 & 1.5 \\
 10 & 100 & 3 & \tau & 66.7 & 11.9 & 12.0 & 7.5 & 4.3 \\
 20 & 400 & 3 & \tau & 96.4 & 7.6 & 23.9 & 8.6 & 1.5 \\
\end{array} \]

Table 2. Power at the 5% level

\[ a = b = -1 \]

\[ a = b = -2 \]

\[ a = -3, b = -1 \]

\[ b = -1 \]

\[ b = -2 \]

\[ b = -3 \]

\[ b = -4 \]

\[ \text{NOTES:} \ k, a \text{ and } b \text{ are such that } p_i = \exp(c_i/N^k T), \text{ where } c_i \sim U(a, b). \text{ See Table 1 for an explanation of the rest.} \]

\(^9\)See Demetrescu and Hanck (2011, sec. 4) for a comprehensive Monte Carlo comparison with their tests.
Table 3. Power at the 5% level

| N  | T  | Case | \( \kappa = 0 \) | \( \kappa = 1/2 \) |
|-----|-----|------|-----------------|-----------------|
|     |     |      | \( \tau \) | \( \hat{\tau}_{IV} \) | \( \hat{S}_{NT} \) | \( \tau_i \) | \( \tau_{IV,i} \) | \( \tau \) | \( \hat{\tau}_{IV} \) | \( \hat{S}_{NT} \) | \( \tau_i \) | \( \tau_{IV,i} \) |
| 10  | 100 | 1    | 97.1 | 93.4 | 36.4 | 15.2 | 5.5 | 53.6 | 7.1 | 12.4 | 7.5 | 2.5 |
| 20  | 400 | 1    | 100.0 | 93.4 | 67.3 | 19.8 | 1.7 | 72.5 | 1.5 | 15.9 | 6.7 | 0.5 |
| 10  | 100 | 2    | 92.6 | 93.4 | 24.9 | 13.0 | 7.9 | 66.3 | 9.1 | 10.4 | 7.4 | 3.9 |
| 20  | 400 | 2    | 100.0 | 93.4 | 50.5 | 17.1 | 3.0 | 92.5 | 2.7 | 15.2 | 6.9 | 1.1 |
| 10  | 100 | 3    | 93.6 | 93.4 | 25.3 | 13.2 | 8.0 | 66.2 | 8.8 | 10.0 | 7.5 | 3.9 |
| 20  | 400 | 3    | 100.0 | 93.4 | 50.0 | 17.1 | 3.0 | 92.4 | 2.7 | 14.9 | 6.9 | 1.1 |
|     |     |      | \( a = b = -5 \) | \( a = b = -10 \) | \( a = -15 \), \( b = -5 \) |
| 10  | 100 | 1    | 99.6 | 93.4 | 66.9 | 27.2 | 12.9 | 88.7 | 20.2 | 23.5 | 10.8 | 3.8 |
| 20  | 400 | 1    | 100.0 | 93.4 | 94.6 | 39.9 | 5.0 | 99.4 | 6.6 | 32.3 | 9.9 | 0.8 |
| 10  | 100 | 2    | 98.5 | 93.4 | 47.7 | 22.0 | 16.6 | 84.4 | 19.8 | 26.0 | 9.5 | 1.6 |
| 20  | 400 | 2    | 100.0 | 93.4 | 82.0 | 32.3 | 7.8 | 99.6 | 8.7 | 26.0 | 9.5 | 1.6 |
| 10  | 100 | 3    | 98.7 | 93.4 | 47.1 | 22.8 | 16.6 | 85.8 | 20.1 | 17.5 | 9.9 | 5.6 |
| 20  | 400 | 3    | 100.0 | 93.4 | 81.9 | 32.8 | 7.8 | 99.6 | 8.8 | 26.1 | 9.6 | 1.6 |
|     |     |      | \( a = -15 \), \( b = -5 \) |
| 10  | 100 | 1    | 99.6 | 93.4 | 65.8 | 25.2 | 13.9 | 82.8 | 19.1 | 23.4 | 10.6 | 3.8 |
| 20  | 400 | 1    | 100.0 | 93.4 | 93.6 | 37.1 | 5.6 | 97.4 | 7.1 | 31.3 | 9.8 | 0.8 |
| 10  | 100 | 2    | 97.1 | 93.4 | 48.0 | 20.2 | 17.0 | 78.3 | 19.9 | 16.5 | 9.3 | 5.8 |
| 20  | 400 | 2    | 100.0 | 93.4 | 82.8 | 28.9 | 8.5 | 97.4 | 9.0 | 24.9 | 9.1 | 1.6 |
| 10  | 100 | 3    | 97.9 | 93.4 | 47.6 | 20.7 | 17.1 | 79.6 | 19.5 | 16.6 | 9.5 | 5.8 |
| 20  | 400 | 3    | 100.0 | 93.4 | 82.8 | 29.3 | 8.5 | 97.4 | 9.2 | 24.8 | 9.1 | 1.7 |

NOTES: See Tables 1 and 2 for an explanation.

should in principle apply to \( \tau_{IV} \). However, this is not what we observe. In fact, on the contrary, unless \( a \) and \( b \) are relatively large in absolute value, the power of this test is actually decreasing in the sample size, which again is probably because \( T < N^3 \). The choice \( \kappa = 1/2 \) is similarly too fast for \( \tau_i \) and \( \tau_{IV,i} \) to have nonnegligible power.

- When \( \kappa = 0 \) the power of the time series tests increases slightly in \( T \), and for the panel tests there is a corresponding power increase in \( N \) and \( T \) when \( \kappa = 1/2 \). For larger sample sizes, however, the power is quite flat in \( N \) and \( T \), which is in accordance with our expectations, since for these combinations of tests and values of \( \kappa \) there should be no dependence on the sample size, at least not asymptotically.

- The best power among the time series tests is obtained by using \( \tau_i \), and among the panel tests \( \tau \) is clearly the preferred choice. Of course, given their size distortions under the null and the fact that the reported powers are not size-adjusted, \( \tau_{IV,i} \) and \( \tau_{IV} \) are actually expected not to perform as well as the new tests. However, the difference in power is way larger than the required compensation for size. In fact, it is not uncommon for the power of the new tests to be more than two times as powerful as the other tests. To take an extreme example, consider the case when \( \kappa = 1/2 \) and \( a = b = -5 \), in which the power of \( \tau \) can be up to 30 times larger than that of \( \tau_{IV} \!).

- Power depends on the nature of the heteroscedasticity. This is particularly clear when \( \kappa = 1/2 \), \( a = -3 \) and \( b = -1 \), in which case the power of \( \tau \) raises from about 20% in case 1 to about 30% in cases 2 and 3, which is just as expected based on the example given in Section 3.2. Unreported results suggest that there is also some variation in power coming from the location of the variance break, which is again in accordance with our expectations.

- The power of \( \tau \) when \( a = b = -2 \) and \( a = b = -10 \) is about the same as when \( a = -3 \) and \( b = -1 \), and \( a = -15 \) and \( b = -5 \), respectively, which confirms the theoretical prediction that the power of this test should be driven mainly by \( \mu_c \). However, there is also a slight tendency for power to decrease as we go from \( a = b = -10 \) to \( a = -15 \) and \( b = -5 \), although the effect vanishes as \( N \) and \( T \) increase. This is in agreement with Theorem 2, suggesting that there should be a second-order effect working through \( \sigma_c^2 \).

The results reported in Tables 1–3 are just a small fraction of the complete set of results that was generated. Some of the results that for space constraints are not reported here, most of which pertains to the selection of the groups, can be found in the supplement. The following summarizes the findings based on those results:

- The QML group selection approach works very well in small samples. The “cost” in terms of test performance of having to estimate the groups is very small. Indeed, the size and power results based on known groups are almost identical to the ones reported in Tables 1–3.

- As pointed out by a referee, when \( N_m = N_{m-1} = 1 \), such that the number of groups is equal to \( N \), then \( \Delta R_{w,t} = \hat{w}_{i,t} \hat{\epsilon}_{i,t} = \text{sign}(\hat{\epsilon}_{i,t}) \), and sign-based tests have been shown to be robust to certain types of heteroscedasticity (see...
Demetrescu and Hanck (2011). Hence, while our theoretical results require that the size of each group goes to infinity with \( N \), the new tests are still expected to perform quite well even when the groups are very small, and this is also what we find.

- While size accuracy is basically unaffected by the size of the groups, as expected, accounting for the group structure can lead to substantial gains in power.

All-in-all, the results reported here and in the supplement lead to us to conclude that the new tests perform well in small samples, and also when compared to some of the existing tests. They should therefore be a valuable addition to the already existing menu of panel unit root tests.

5. CONCLUDING REMARKS

This article focuses on the problem of how to test for unit roots in panels where the errors are contaminated by time series and cross-section dependence, and in addition general unconditional heteroscedasticity. In the article, we assume that the heteroscedasticity is deterministic, as when considering modulated heteroscedasticity. In the article, we assume that the roots in panels where the errors are contaminated by time series distortions than the tests of Demetrescu and Hanck (2011) and, at the same time, have much higher power. These findings suggest that they should be a useful addition to the existing menu of unit root tests.

APPENDIX: PROOFS

Lemma A.1. Under Assumptions 1–3,

\[
\Delta R_{wit} = \Delta R_{wit} + g_{it} + O_p \left( \frac{1}{C_{2NT}} \right),
\]

\[
\frac{1}{\sqrt{T}} \hat{R}_{wit} = \frac{1}{\sqrt{T}} \left( R_{wit} + G_{it} \right) + O_p \left( \frac{1}{C_{2NT}} \right),
\]

where

\[
\Delta R_{wit} = \epsilon_{i,t} + (\rho_i - 1) w_{m,t} R_{it-1},
\]

\[
R_{wit} = \sum_{s=p+2}^t \epsilon_{i,s} + (\rho_i - 1) \sum_{s=p+2}^t w_{m,s} R_{it-1},
\]

\[
g_{it} = \Theta_i^t H^{-1} w_{m,t} v_i - \frac{1}{2} w_{m,t}^3 (\bar{\epsilon}_{0m,t} + \bar{\epsilon}_{2m,t} + \bar{\epsilon}_{6m,t}) \epsilon_{i,t},
\]

\[
G_{it} = \Theta_i^t H^{-1} \sum_{s=p+2}^t w_{m,s} v_s - \frac{1}{2} \sum_{s=p+2}^t w_{m,s}^3 \bar{\epsilon}_{0m,t} \epsilon_{i,t},
\]

\[
\tau_{xm,t} = \frac{1}{(N - N_m)} \sum_{i=N_m+1}^{N} c_{xi,t},
\]

with \( i = N_m+1, \ldots, N_m \), \( R_{it} = \phi_i(L) u_{i,t} \), \( c_{it} = (\epsilon_{it}^2 - \sigma_m^2) \epsilon_{it} \), \( c_{it}^2 = 2 \epsilon_{it} a_{it} \), \( c_{it}^3 = -2 (\hat{\phi}_i - \phi_i) \epsilon_{it} \Delta x_{it-1} \), \( a_{it} = \Theta_i^t H^{-1} v_i - d_i \hat{F}_t \), \( v_i = (\hat{F}_t - H \hat{F}_t) \), \( d_i = (\hat{\phi}_i - (H^{-1}) \phi_i) \), \( H \) is a \( r \times r \) positive definite matrix and \( C_{2NT} = \min(N, \sqrt{T}) \).

Proof of Lemma A.1. Let \( x_{it} = (y_{i,t}, \ldots, y_{i,t-p+1})' \) and \( \Phi_i = (\phi_{i1}, \ldots, \phi_{ip})' \). We have

\[
R_{it} = \phi_i(L) u_{i,t} = \phi_i(L) (y_{i,t} - \alpha_i) = y_{i,t} - \hat{\Phi}_i x_{i,t-1} - \phi_i(1) \alpha_i,
\]

whose first difference is given by

\[
\Delta R_{it} = \phi_i(L) \Delta y_{it} = \Delta y_{it} - \hat{\Phi}_i' \Delta x_{i,t-1}.
\]

It follows that

\[
\Delta \hat{R}_{it} = (\rho_i - 1) R_{it-1} + \epsilon_{i,t}, \quad (A.1)
\]

where \( \epsilon_{i,t} \) is defined in Equation (4).

Consider \( \Delta \hat{R}_{it} \), which we can expand as

\[
\Delta \hat{R}_{it} = \Delta y_{it} - \hat{\Phi}_i' \Delta x_{i,t-1} = \Delta R_{it} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1}.
\]

Note that by Taylor expansion of \( \exp(x) \) about \( x = 0 \), \( \exp(x) = 1 + x + o(1) \), suggesting that \( \rho_i = \exp(c_i/N^s T) = 1 + c_i/N^s T + o_p(1) \). This implies

\[
\frac{1}{\sqrt{T}} \sum_{t=p+2}^T \Delta x_{i,t-1} \Delta \hat{R}_{it} = \frac{c_i}{N^s T^{3/2}} \times \sum_{t=p+2}^T \Delta x_{i,t-1} R_{it-1} + \frac{1}{\sqrt{T}} \sum_{t=p+2}^T \Delta x_{i,t-1} \epsilon_{i,t} + o_p(1).
\]

In the proof of Theorem 1, we show that \( R_{it-1}/\sqrt{T} \) satisfies an invariance principle, and therefore that the first term on the right-hand side is \( O_p(1/N^s \sqrt{T}) \). As for the second term, by the Wold decomposition, \( 1/\phi(L) \) can be expanded as \( \phi^*(L) = 1/\phi(L) = \sum_{j=0}^{\infty} \phi_0^* L^j \) with \( \phi_0^* = 1. \) It follows that \( \Delta y_{it} = (\rho_i - 1) (y_{it} - \alpha_i) + v_{it} \), where \( v_{it} = \phi_i'(L) \epsilon_{i,t} \). Hence, since \( e_{i,t} \) is serially uncorrelated, we have that \( E(\epsilon_{i,t} \Delta x_{i,t-1}) = 0 \) for all \( i, k \) and \( s \). The variance of \( \sum_{t=p+2}^T \Delta x_{i,t-1} \epsilon_{i,t} \) is bounded. Hence, \( \sum_{t=p+2}^T \Delta x_{i,t-1} \epsilon_{i,t} / \sqrt{T} = O_p(1) \). Subsequently, using \( \hat{\Phi}_i \) to
denote the OLS estimator of $\Phi_i$,

$$\sqrt{T}(\hat{\Phi}_i - \Phi_i) = \left( \frac{1}{T} \sum_{t=p+1}^{T} \Delta x_{i,t-1}(\Delta x_{i,t-1})' \right)^{-1} \times \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \Delta x_{i,t-1} \Delta R_{i,t}$$

$$= \left( \frac{1}{T} \sum_{t=p+1}^{T} \Delta x_{i,t-1}(\Delta x_{i,t-1})' \right)^{-1} \times \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \Delta x_{i,t-1} \epsilon_{i,t} + o_p(1) = O_p(1).$$

The presence of the factors therefore does not affect the rate of consistency of $\hat{\Phi}_i$. Making use of this result, we obtain

$$\Delta \hat{R}_{i,t} = \Delta y_{i,t} - \hat{\Phi}_i \Delta x_{i,t-1} = \Delta R_{i,t} - (\hat{\Phi}_i - \Phi_i) \Delta x_{i,t-1}$$

$$= \Delta R_{i,t} + O_p \left( \frac{1}{\sqrt{T}} \right).$$

and therefore, with $R_{i,t-1} = O_p(\sqrt{T})$ and $(\rho_i - 1) = O_p(1/N^2 T)$,

$$\Delta \hat{R}_{i,t} = \Delta R_{i,t} + O_p \left( \frac{1}{\sqrt{T}} \right) = \epsilon_{i,t} + (\rho_i - 1) R_{i,t-1}$$

$$+ O_p \left( \frac{1}{\sqrt{T}} \right) = \epsilon_{i,t} + O_p \left( \frac{1}{\sqrt{T}} \right).$$

Since $\epsilon_{i,t} = \Theta_i \epsilon_{i,t} + \epsilon_{i,t}$, we have that the error incurred by applying the principal components method to $\Delta \hat{R}_{i,t}$ rather than to $\epsilon_{i,t}$ is negligible (a detailed proof is available upon request). To also show that the estimation is unaffected by the heteroscedasticity in $\epsilon_{i,t}$, we verify Assumption C in Bai (2003). Toward this end, note that, since $N_m = \lfloor \lambda_m N \rfloor$,

$$\frac{1}{N} \sum_{i=1}^{N} E(\epsilon_{i,t}^2)$$

$$= \sum_{m=1}^{M+1} \sum_{i=N_{m-1}+1}^{N} E(\epsilon_{i,t}^2)$$

$$= \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \frac{1}{(N_m - N_{m-1})} \sum_{i=N_{m-1}+1}^{N} E(\epsilon_{i,t}^2)$$

$$= \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \sigma^2_{m,t},$$

which is bounded under our assumptions, and therefore so is $\sum_{i=1}^{N} E(\epsilon_{i,t}^2)/N = \sum_{i=1}^{N} \epsilon_{i,t}^2/N$. Similarly,

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} E(\epsilon_{i,t}\epsilon_{j,t})$$

$$= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E(\epsilon_{i,t}\epsilon_{j,t}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \sigma^2_{m,t},$$

which is bounded too, because $\sum_{t=1}^{T} \sigma^2_{m,t}/T \to \int_{r=0}^{1} \sigma^2_{m,t}(r)dr < \infty$ as $T \to \infty$. Similarly, since $\int_{r=0}^{1} \Sigma(r)dr > 0$, $\sum_{t=1}^{T} F_tF_t'/T$ converges to a positive definite matrix. These results imply that Assumption C of Bai (2003) holds and therefore that the principal component estimators $\hat{\Theta}_i$ and $\hat{F}_i$ of $\Theta_i$ and $F_i$, respectively, are unaffected by the heteroscedasticity.

Let us now consider $\Delta \hat{R}_{i,t} = \hat{\omega}_{i,t} \hat{\epsilon}_{i,t} = \hat{\epsilon}_{i,t} / \hat{\sigma}_{m,t}$. Here,

$$\hat{\epsilon}_{i,t} = \epsilon_{i,t} + (\rho_i - 1) R_{i,t-1} + a_{i,t}$$

$$- (\hat{\Phi}_i - \Phi_i) \Delta x_{i,t-1},$$

where, following Bai and Ng (2004, p. 1154),

$$a_{i,t} = \Theta_i F_i - \Theta_i \hat{F}_i = \Theta_i H^{-1}(\hat{F}_i - H F_i) - (\hat{\Theta}_i - (H^{-1})^\prime \Theta_i)^\prime \hat{F}_i = \Theta_i H^{-1} v_t - d_i \hat{F}_i,$$

with $v_t$ and $d_i$ implicitly defined. Hence,

$$\hat{\epsilon}_{i,t}^2 = (\epsilon_{i,t} + (\rho_i - 1) R_{i,t-1} + a_{i,t} - (\hat{\Phi}_i - \Phi_i) \Delta x_{i,t-1})^2$$

$$= \hat{\epsilon}_{i,t}^2 + c_{i,t} + \cdots + c_{9i,t},$$

with $c_{i,t} = (\rho_i - 1)^2 R_{i,t-1}^2$,

$$c_{2i,t} = a_{i,t}^2$$

$$c_{3i,t} = (\hat{\Phi}_i - \Phi_i) \Delta x_{i,t-1} \Delta x_{i,t-1} (\hat{\Phi}_i - \Phi_i)$$

$$c_{4i,t} = 2 \epsilon_{i,t} a_{i,t}$$

$$c_{5i,t} = 2 (\rho_i - 1) \epsilon_{i,t} R_{i,t-1}$$

$$c_{6i,t} = -2 (\hat{\Phi}_i - \Phi_i) \epsilon_{i,t} \Delta x_{i,t-1} - (\hat{\Phi}_i - \Phi_i) \Delta x_{i,t-1}$$

$$c_{7i,t} = 2 (\rho_i - 1) a_{i,t} R_{i,t-1}$$

$$c_{8i,t} = -2 (\hat{\Phi}_i - \Phi_i) a_{i,t} \Delta x_{i,t-1}$$

$$c_{9i,t} = -2 (\rho_i - 1) (\hat{\Phi}_i - \Phi_i) R_{i,t-1} \Delta x_{i,t-1}.$$

For ease of notation and without loss of generality, in what follows we focus on the first group containing the first $N_1$ cross-section units. Let us therefore consider $\hat{\sigma}_{1,t}^2$, which is $\sigma_{1,t}^2$ for the first group. Letting $c_{01,t} = (\epsilon_{i,t} - \sigma_{1,t}^2)$ and using $c_{11,t} = \sum_{i=1}^{N_1} c_{i,t}/N_1$ to denote the average $c_{i,t}$ for this group, we have

$$\hat{\sigma}_{1,t}^2 = \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t}^2 = \frac{1}{N_1} \sum_{i=1}^{N_1} c_{i,t}^2 + c_{11,t} + \cdots + c_{91,t}$$

$$= \sigma_{1,t}^2 + \bar{c}_{1,t} + \cdots + \bar{c}_{91,t} = \sigma_{1,t}^2 + \bar{C}_{1,t},$$

with an obvious definition of $\bar{C}_{1,t}$. This result, together with a second-order Taylor expansion of the inverse square root of $\hat{\sigma}_{1,t}^2$, yields

$$\hat{\sigma}_{1,t} = \frac{1}{\hat{\sigma}_{1,t}}$$

$$= w_{1,t} - \frac{1}{2} w_{1,t}^3 (\hat{\sigma}_{1,t}^2 - \sigma_{1,t}^2) + \frac{3}{4} w_{1,t}^5 (\hat{\sigma}_{1,t}^2 - \sigma_{1,t}^2)^2$$

$$+ O_p \left( \frac{(\hat{\sigma}_{1,t}^2 - \sigma_{1,t}^2)^3}{\hat{\sigma}_{1,t}^2} \right)$$

$$= w_{1,t} - \frac{1}{2} w_{1,t}^3 \bar{C}_{1,t} + \frac{3}{4} w_{1,t}^5 \bar{C}_{1,t}^2 + O_p(\bar{C}_{1,t}).$$

(A.7)
To work out the order of the remainder, we are going to make use of Lemma 1 of Bai and Ng (2004), which states that \( |u_i| = O_p(1/C_{1NT}) \) and \( |d_i| = O_p(1/C_{2NT}) \), where \( C_{1NT} = \min(\sqrt{N}, T) \) and \( C_{2NT} = \min(N, \sqrt{T}) \).

**Remark A.1.** The expansion in Equation (A.7) has to be of second order because, otherwise the resulting remainder in the numerator of \( \tau \), \( \sum_{i=1}^{N} \sum_{i=2}^{T} \Delta R_{w_{i,t}}R_{w_{i,t-1}}/\sqrt{NT} \), would not necessarily be negligible. For the numerator of \( \tau \), \( \sum_{i=1}^{T} \Delta R_{w_{i,t}}R_{w_{i,t-1}}/T \), it is enough with a first-order expansion.

We begin with \( \bar{c}_{01,t} \), whose order is given by

\[
\bar{c}_{01,t} = \frac{1}{N_1} \sum_{i=1}^{N_1} (\epsilon_{i,t}^2 - \sigma_{i,t}^2) = O_p \left( \frac{1}{\sqrt{N}} \right),
\]

as follows from using \( E(\epsilon_{i,t}^2 - \sigma_{i,t}^2) = 0 \), cross-section independence and \( N_1 = \lambda_1 N = O(N) \). As for \( \bar{c}_{21,t} \), by using \( (a^2 + b^2) \leq 2(a^2 + b^2) \) and \( \|AB\| \leq \|A\|\|B\| \) for arbitrary scalars \( a \) and \( b \), and conformable matrices \( A \) and \( B \), it is clear that \( c_{21,t} = a_{i,t}^2 \leq 2(\Theta^{-1}||H^{-1}||^2 ||v_i||^2 + 2||d_i||^2 ||\hat{F}_t||^2 \), and therefore

\[
\bar{c}_{21,t} = \frac{1}{N_1} \sum_{i=1}^{N_1} a_{i,t}^2 \leq \frac{2}{N_1} \sum_{i=1}^{N_1} \|\Theta^{-1}H^{-1}\||^2 ||v_i||^2 \\
+ \frac{2}{N_1} \sum_{i=1}^{N_1} ||d_i||^2 ||\hat{F}_t||^2 = O_p \left( \frac{1}{C_{2NT}} \right),
\]

where \( C_{NT} = C_{1NT} + C_{2NT} = \min(\sqrt{N}, \sqrt{T}) \). Next, consider \( \bar{c}_{41,t} \). Because \( \hat{F}_t = H F_t + \psi_t \), we have \( \epsilon_{i,t}^2 \hat{F}_t = \epsilon_{i,t}^2 H F_t + \epsilon_{i,t}^2 \psi_t \), where, by the Cauchy–Schwarz inequality,

\[
\left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} d_i \right| \leq \left( \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t}^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||d_i||^2 \right)^{1/2} = O_p \left( \frac{1}{C_{2NT}} \right).
\]

It follows that

\[
\left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} d_i^2 H F_t \right| \leq \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} d_i \right| ||H F_t|| = O_p \left( \frac{1}{C_{2NT}} \right),
\]

\[
\left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} d_i^2 \psi_t \right| \leq \left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} d_i \right| ||\psi_t|| = O_p \left( \frac{1}{C_{1NT}C_{2NT}} \right),
\]

and so, via the cross-section independence of \( \epsilon_{i,t} \),

\[
\bar{c}_{41,t} = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} a_{i,t} \right| \leq \left| \frac{2}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} \Theta_i \right| ||H^{-1}|| ||v_i|| \\
+ \left| \frac{2}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,t} d_i \right| ||\hat{F}_t|| = O_p \left( \frac{1}{\sqrt{NC_{1NT}}} \right) + O_p \left( \frac{1}{C_{2NT}} \right).
\]

Similarly, since \( (\rho_i - 1) = O_p(1/NT) \),

\[
\bar{c}_{71,t} = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} (\rho_i - 1)a_{i,t} R_{i,t-1} \right| \leq 2 \left( \frac{1}{N_1} \sum_{i=1}^{N_1} a_{i,t}^2 \right)^{1/2} \times \left( \frac{1}{N_1} \sum_{i=1}^{N_1} (\rho_i - 1)^2 R_{i,t-1}^2 \right)^{1/2} = O_p \left( \frac{1}{\sqrt{N^3 C_{NT}}} \right),
\]

\[
\bar{c}_{81,t} = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} (\Phi_i - \Phi F_t) a_{i,t} \Delta x_{i,t-1} \right| \leq 2 \left( \frac{1}{N_1} \sum_{i=1}^{N_1} a_{i,t}^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||\Phi_i - \Phi F_t||^2 ||\Delta x_{i,t-1}||^2 \right)^{1/2} = O_p \left( \frac{1}{\sqrt{T C_{NT}}} \right).
\]

The same argument can be used to show that \( \bar{c}_{11,t}, \bar{c}_{31,t} \) and \( \bar{c}_{91,t} \) are \( O_p(1/T) \). It remains to consider \( \bar{c}_{51,t} \) and \( \bar{c}_{61,t} \). The orders of these are as follows:

\[
\bar{c}_{51,t} = \frac{2}{N_1} \sum_{i=1}^{N_1} (\rho_i - 1) e_{i,t} R_{i,t-1} \\
= \frac{2}{N_1 N^2 T} \sum_{i=1}^{N_1} c_i e_{i,t} R_{i,t-1} = O_p \left( \frac{1}{N^{1/2} + \sqrt{T}} \right),
\]

\[
\bar{c}_{61,t} = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} (\Phi_i - \Phi) \psi_{i,t} \Delta x_{i,t-1} \right| \leq 2 \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||\Phi_i - \Phi||^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||\psi_{i,t} \Delta x_{i,t-1}||^2 \right)^{1/2} = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

As the above calculations make clear, \( \bar{c}_{01,t}, \bar{c}_{21,t}, \bar{c}_{41,t}, \bar{c}_{71,t}, \bar{c}_{81,t} \) are the leading terms in \( \bar{c}_{1,t} \). Hence, by direct substitution of Equations (A.8)–(A.12) into the definition of \( \bar{c}_{1,t} \),

\[
\bar{c}_{1,t} = \bar{c}_{01,t} + \bar{c}_{41,t} + \bar{c}_{61,t} + O_p \left( \frac{1}{C_{2NT}} \right),
\]

where the order of the remainder follows from the fact that

\[
O_p \left( \frac{1}{C_{1NT}} \right) + O_p \left( \frac{1}{\sqrt{N} C_{1NT}} \right) + O_p \left( \frac{1}{\sqrt{T} C_{NT}} \right) = O_p \left( \frac{1}{C_{2NT}} \right).
\]

Because \( \bar{c}_{01,t} + \bar{c}_{41,t} + \bar{c}_{61,t} = O_p(1/C_{2NT}) \), the remainder term in Equation (A.7) is \( O_p(1/C_{2NT}) \), and therefore

\[
\hat{w}_{1,t} = w_{1,t} - \frac{1}{2} \hat{w}_{1,t} \hat{c}_{1,t} + \frac{3}{4} \hat{w}_{1,t} \hat{c}_{1,t} + O_p \left( \frac{1}{C_{3NT}} \right).
\]
The result in Equation (A.16) suggests that \( \Delta \hat{R}_{i,t} \) can be expanded as

\[
\Delta \hat{R}_{ui,t} = \hat{w}_{1,t} \hat{\epsilon}_{i,t} + (\rho - 1)R_{i,t-1} + a_{i,t} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1}
\]

\[
= \Delta R_{ui,t} + w_{1,t}(a_{i,t} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1})
\]

\[
- \frac{1}{2} w_{1,t}^3 C_{1,t} \Delta R_{ui,t}
\]

\[
+ \frac{3}{4} w_{1,t}^5 C_{1,t}^2 \Delta R_{ui,t}
\]

\[
+ \frac{3}{4} w_{1,t}^6 C_{1,t}^2 (a_{i,t} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1}) + O_p \left( \frac{1}{C_{3NT}} \right),
\]

(A.17)

where \( \Delta R_{ui,t} = \epsilon_{i,t} + (\rho - 1)w_{1,t}R_{i,t-1} \). Consider the second term on the right-hand side, \( w_{1,t}(a_{i,t} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1}) \). By using \( a_{i,t} = \Theta_i H^{-1} v_t - d_i \bar{F}_t = \Theta_i H^{-1} v_t + O_p(1/C_{2NT}) \) and \( |(\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1}| \leq ||\hat{\Phi}_i - \Phi_i|| ||\Delta x_{i,t-1}|| = O_p(1/\sqrt{T}) = O_p(1/C_{2NT}) \), it is clear that

\[
w_{1,t}(a_{i,t} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1}) = \Theta_i H^{-1} w_{1,t} v_t + O_p \left( \frac{1}{C_{2NT}} \right).
\]

The order of \( \Delta R_{ui,t} \) is equal to that of \( C_{1,t} \), and therefore

\[
\Delta R_{ui,t} = (\zeta_{01,t} + \zeta_{41,t} + \zeta_{61,t}) \Delta R_{ui,t} + O_p \left( \frac{1}{C_{2NT}} \right).
\]

Similarly, since \( C_{1,t} \) and \( a_{i,t} \) are both \( O_p(1/C_{NT}) \),

\[
|\zeta_{1,t}((a_{i,t} - ||\hat{\Phi}_i - \Phi_i|| ||\Delta x_{i,t-1}||))|
\]

\[
\leq O_p \left( \frac{1}{C_{3NT}} \right) + O_p \left( \frac{1}{\sqrt{T}C_{NT}} \right),
\]

which is clearly \( O_p(1/C_{2NT}) \). The other terms are all of higher order than this. It follows that

\[
\Delta \hat{R}_{ui,t} = \Delta R_{ui,t} + \Theta_i H^{-1} w_{1,t} v_t - \frac{1}{2} w_{1,t}^3 (\zeta_{01,t} + \zeta_{41,t} + \zeta_{61,t})
\]

\[
\times \Delta R_{ui,t} + O_p \left( \frac{1}{C_{2NT}} \right)
\]

\[
= \Delta R_{ui,t} + \Theta_i H^{-1} w_{1,t} v_t - \frac{1}{2} w_{1,t}^3 (\zeta_{01,t} + \zeta_{41,t} + \zeta_{61,t}) \epsilon_{i,t}
\]

\[
+ O_p \left( \frac{1}{C_{2NT}} \right)
\]

\[
= \Delta R_{ui,t} + g_{i,t} + O_p \left( \frac{1}{C_{2NT}} \right),
\]

(A.18)

with an obvious definition of \( g_{i,t} \), and where the second equality uses the approximation \( \Delta R_{ui,t} = \epsilon_{i,t} + (\rho - 1)w_{1,t}R_{i,t-1} = \epsilon_{i,t} + o_p(1) \), which is valid in the present case (details are available upon request). This establishes the first of the two required results.

The second result requires more work. Note first that by using Equation (A.16),

\[
\frac{1}{\sqrt{T}} \hat{\epsilon}_{i,t} = \frac{1}{\sqrt{T}} \sum_{s=p+2}^t \Delta \hat{R}_{ui,s}
\]

\[
= \frac{1}{\sqrt{T}} R_{ui,t} + \frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s}(a_{i,s} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,s-1})
\]

\[
- \frac{1}{2 \sqrt{T}} \sum_{s=p+2}^t w_{1,s}^3 \Delta R_{ui,s}
\]

\[
- \frac{1}{2 \sqrt{T}} \sum_{s=p+2}^t w_{1,s}^5 \Delta R_{ui,s}
\]

\[
+ \frac{3}{4 \sqrt{T}} \sum_{s=p+2}^t w_{1,s}^7 \Delta R_{ui,s}
\]

\[
+ \frac{3}{4 \sqrt{T}} \sum_{s=p+2}^t w_{1,s}^9 \Delta R_{ui,s}
\]

\[
+ O_p \left( \frac{\sqrt{T}}{C_{NT}} \right),
\]

(A.19)

where again all terms that are \( O_p(1/C_{2NT}) \) can be treated as negligible.

Consider \( \sum_{s=p+2}^t w_{1,s}(a_{i,s} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,s-1})/\sqrt{T} \). Via \( \hat{F}_s = HF_s + v_s \), and then leading term approximation, \( \sum_{s=p+2}^t \hat{F}_s/\sqrt{T} = \sum_{s=p+2}^t w_{1,s}HF_s/\sqrt{T} + O_p(1) = O_p(1) \), implying

\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s} a_{i,s}
\]

\[
= \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s} v_s - d_i \frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s} \hat{F}_s
\]

\[
= \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s} v_s + O_p \left( \frac{1}{C_{2NT}} \right),
\]

which, together with \( (\hat{\Phi}_i - \Phi_i)' \sum_{s=p+2}^t w_{1,s} \Delta x_{i,s-1} - 1/\sqrt{T} = O_p(1/\sqrt{T}) \), yields

\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s}(a_{i,s} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,s-1}) = \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p+2}^t w_{1,s} v_s + O_p \left( \frac{1}{C_{2NT}} \right).
\]

(A.20)
Next, consider
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{1,s} \Delta R_{w,s}
\]
where \( \sum_{s=p+2}^{t} w^3_{s,s} c_{1,s} \Delta R_{w,s}/\sqrt{T} \) is again not of sufficiently low order to be ignored. Let us therefore consider \( \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{1,s} \Delta R_{w,s}/\sqrt{T} \). Because \( \epsilon_{i,s} \) is serially uncorrelated and independent of \( F_t \), we have \( E(\epsilon_{i,s} \Delta x_{k,s-1}) = 0 \). Moreover, \( \sum_{s=p+2}^{t} w^3_{k,s} \epsilon_{i,s} \Delta x_{k,s-1}/\sqrt{T} = O_p(1) \), implying
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{1,s} \epsilon_{i,s}
\]
\[
= \frac{1}{N_1} \sum_{k=1}^{N_1} \Phi_k - \Phi_k \left( \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^2_{k,s} \epsilon_{i,s} \Delta x_{k,s-1} \epsilon_{i,s} \right)
\]
\[
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \| \Phi_k - \Phi_k \| \right)^{1/2}
\times \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left\| \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^2_{k,s} \epsilon_{i,s} \bar{c}_{1,s} \Delta x_{k,s-1} \right\| \right)^{1/2}
\]
\[
= O_p \left( \frac{1}{T} \right).
\]
Similarly, since \( \sum_{s=p+2}^{t} w^3_{k,s} \epsilon_{i,s} \Delta x_{k,s-1} R_{t,s-1}/T^{3/2} = O_p(1) \), we can show
\[
(\rho - 1) \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^4_{k,s} \bar{c}_{61,s} \Delta R_{w,s}
\]
\[
= \frac{1}{N_1 N_1} \sum_{k=1}^{N_1} (\Phi_k - \Phi_k) \frac{1}{T^{3/2}} \sum_{s=p+2}^{t} w^3_{k,s} \epsilon_{i,s} \Delta x_{k,s-1} R_{t,s-1}
\]
\[
= O_p \left( \frac{1}{N_1 T} \right),
\]
and therefore
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{61,s} \Delta R_{w,s}
\]
\[
= \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} \left\| \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \epsilon_{i,s} d_k^T HF_{i,s} \Delta R_{w,s} \right\|
\]
\[
= \frac{1}{N_1} \sum_{k=1}^{N_1} d_k^T \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} HF_{i,s} w^2_{k,s} \epsilon_{i,s}
\]
\[
+ \frac{c_i}{N_1} \sum_{k=1}^{N_1} d_k^T \frac{1}{T^{3/2}} \sum_{s=p+2}^{t} \epsilon_{i,s} w^3_{k,s} HF_{i,s} R_{t,s-1}
\]
\[
= O_p \left( \frac{1}{C_{2NT}} \right).
\]
Hence, since \( -2 \sum_{k=1}^{N_1} \epsilon_{k,s} d_k' HF_{i,s} / N_1 \) is the leading term in \( \bar{c}_{41,s} \),
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{41,s} \Delta R_{w,s}
\]
\[
= - \frac{2}{\sqrt{T} N_1} \sum_{k=1}^{N_1} \sum_{s=p+2}^{t} w^3_{k,s} \epsilon_{k,s} d_k' HF_{i,s} \Delta R_{w,s} + o_p(1)
\]
\[
= O_p \left( \frac{1}{C_{2NT}} \right).
\]
Thus, while nonnegligible in \( \bar{c}_{41,s} \) and \( \bar{c}_{61,s} \), are in fact negligible in \( \bar{c}_{41,s} \) and \( \bar{c}_{61,s} \), will generally be dominated by that of \( \epsilon_{i,s} \). Hence, unless otherwise stated, in what follows we are going to work with the approximation \( \Delta R_{w,s} = \epsilon_{i,s} + o_p(1) \).
Both \( R_{i,s}/\sqrt{T} \) and \( R_{w,s}/\sqrt{T} \) satisfy an invariance principle. By using this and the fact that \( (\rho - 1) \sim O_p(1/N^T) \),
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{1,s} \Delta R_{w,s}
\]
\[
= \frac{1}{\sqrt{T} N_1} \sum_{k=1}^{N_1} \sum_{s=p+2}^{t} w^3_{k,s} (\rho - 1) R^2_{k,s-1} \Delta R_{w,s}
\]
\[
= \frac{c_i}{T^{3/2} N^2 \sqrt{T_1}} \sum_{k=1}^{N_1} \sum_{s=p+2}^{t} w^3_{k,s} \Delta R_{w,s} = O_p \left( \frac{1}{T} \right).
\]
It is also not difficult to show that
\[
\left| \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w^3_{k,s} \bar{c}_{31,s} \Delta R_{w,s} \right|
\]
\[
\leq \left| \frac{1}{\sqrt{T} N_1} \sum_{k=1}^{N_1} \sum_{s=p+2}^{t} w^3_{k,s} (\Phi_k - \Phi_k) \Delta x_{k,s-1} \Delta x_{k,s-1}^T \right|
\times \left( \Phi_k - \Phi_k \right) \Delta R_{w,s}
\]
\[
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left\| \Phi_k - \Phi_k \| \right\|^2 \right)^{1/2}
\times \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left\| \Phi_k - \Phi_k \| \right\|^4 \right)^{1/2}
\]
\[
= O_p \left( \frac{1}{T} \right).
\]
and similarly,
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 c_{51,s} \Delta R_{wi,s} = \frac{1}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\rho_k - 1) \epsilon_{k,s} R_{k,s-1} \Delta R_{wi,s} = O_p\left( \frac{1}{N^{k+1/2}T} \right).
\]

\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 c_{91,s} \Delta R_{wi,s} = -\frac{2}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\rho_k - 1) (\Phi_k - \Phi_k') R_{k,s-1} \Delta R_{wi,s} = O_p\left( \frac{1}{N^{k+1/2}T} \right),
\]

which are all \(O_p(1/C_{2NT})\).

The terms involving \(a_{i,t}\) can be evaluated in a similar fashion. In particular, we have
\[
\left| \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 c_{71,s} \Delta R_{wi,s} \right| = \left| \frac{2}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\rho_k - 1) a_{k,s} R_{k,s-1} \Delta R_{wi,s} \right| \leq 2 \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| \frac{c_k}{N^kT^{3/2}} \right| \sum_{s=p+2}^{t} w_{1,s}^3 R_{k,s-1} \Delta R_{wi,s} \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| a_{k,s} \right|^2 \right)^{1/2} = O_p\left( \frac{1}{N^{k+1/2}T^{3/2}} \right).
\]

where we have made use of the fact that \(\sum_{s=p+2}^{t} w_{1,s}^3 R_{k,s-1} \Delta R_{wi,s} / T = O_p(1)\) and the results of Bai and Ng (2004, p. 1158), from which we deduce that \(\sum_{s=p+2}^{t} a_{k,s} / \sqrt{T} = O_p(1/C_{2NT})\). Clearly, \(O_p(1/N^{k+1/2}T^{3/2}C_{2NT}) = O_p(1/C_{2NT})\).

Moreover, since \(c_{2,t} = a_{k,s} = (\Theta_k' H^{-1} v_s) + (dF')^2\),
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 c_{21,s} \Delta R_{wi,s} = \frac{1}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 a_{k,s}^2 R_{wi,s} = O_p\left( \frac{1}{N^{k+1/2}T^{3/2}} \right).\]

Now, since \(\Theta_k' H^{-1} v_s\) and \(dF'\) are just scalars, \(\Theta_k' H^{-1} v_s dF' = \tilde{F}_t v_s (H^{-1})' \Theta_k dF'\). Hence, since \(||a + b||^2 \leq ||a||^2 + ||b||^2||, \(\sum_{s=p+2}^{t} w_{1,s}^3 (\tilde{F}_t v_s) \Delta R_{wi,s} v_s^2 \leq \sum_{s=p+2}^{t} w_{1,s}^3 (\Delta R_{wi,s})^2 ||\tilde{F}_t||^2 ||v_s||^2\), suggesting that
\[
\left| \frac{1}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 \Theta_k' H^{-1} v_s dF' \tilde{F}_t \Delta R_{wi,s} \right| \leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \sum_{s=p+2}^{t} w_{1,s}^3 (\Delta R_{wi,s})^2 ||\tilde{F}_t||^2 ||v_s||^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||\Theta_k||^2 ||dF'||^2 \right)^{1/2} H^{-1} ||1|| = O_p\left( \frac{1}{C_{2NT}} \right).
\]

Also, by the cyclical property of the trace, \(\text{tr}(dF') = \text{tr}(dF' \tilde{F}_t v_s) = \text{tr}(dF' \tilde{F}_t F_{s}').\)

Therefore, since \(\Delta R_{wi,s} = \epsilon_{i,t} + o_p(1)\), where \(\epsilon_{i,t}\) is asymptotically uncorrelated with \(\tilde{F}_t\),
\[
\frac{1}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (dF')^2 \Delta R_{wi,s} = \text{tr} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} dF' \frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 \tilde{F}_t F_{i,s}' \epsilon_{i,t} \right) = O_p\left( \frac{1}{C_{2NT}} \right),
\]

suggesting
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 c_{21,s} \Delta R_{wi,s} = \frac{1}{\sqrt{T}N_1} \sum_{s=p+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\Theta_k' H^{-1} v_s)^2 \Delta R_{wi,s} + O_p\left( \frac{1}{C_{2NT}} \right).
\]

As for the remaining term, according to Bai (2003, p. 166), \(v_s = H' \Lambda^{-1} \sum_{k=1}^{N} \Theta_{k} \epsilon_{k,s} / N + o_p(1)\), where \(\Lambda = \lim_{N \to \infty} \sum_{k=1}^{N} \Theta_{k} \Theta_{k}' / N\). By using this and the fact that \(\epsilon_{i,t}\) is cross-section independent, we obtain
\[
\frac{1}{\sqrt{T}} \sum_{s=p+2}^{t} w_{1,s}^3 v_s \epsilon_{i,t} \epsilon_{i,t} \epsilon_{i,t} \epsilon_{i,t} = H' \Lambda^{-1} \frac{1}{\sqrt{T}N^2} \times \sum_{s=p+2}^{t} \sum_{k=1}^{N} \sum_{l=1}^{N} \Theta_{k} w_{1,s}^3 \epsilon_{k,s} \epsilon_{k,s} \Theta_{l} (\Lambda^{-1})' H + o_p(1)\]
\[ H'\Lambda^{-1} \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N} \Theta_k w_{1,s}^{i} \varepsilon_{i,s}^{2} \Theta_k' (\Lambda^{-1})' H + o_p(1) \]

\[ = H'\Lambda^{-1} \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \Theta_k w_{1,s}^{i} \varepsilon_{i,s}^{3} \Theta_k' (\Lambda^{-1})' H + H'\Lambda^{-1} \times \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k \neq i} \Theta_k w_{1,s}^{i} \varepsilon_{i,s}^{2} \Theta_k' (\Lambda^{-1})' H + o_p(1) \]

\[ = H'\Lambda^{-1} \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \Theta_k w_{1,s}^{i} \varepsilon_{i,s}^{3} \Theta_k' (\Lambda^{-1})' H + H'\Lambda^{-1} \sum_{s=p_t+2}^{t} \Lambda_i w_{1,s}^{3} \varepsilon_{i,s} (\Lambda^{-1})' H + o_p(1) \]

\[ = O_p \left( \frac{\sqrt{T}}{N^2} \right) + O_p \left( \frac{1}{N} \right) = O_p \left( \frac{\sqrt{T}}{C_{NT}} \right) \]

where \( \Lambda_i = \lim_{N \to \infty} \sum_{k=1}^{N} \sigma_k^2 \Theta_k' \Theta_k / N \) such that \( \Lambda_i T \rightarrow \Lambda(t) > 0 \) as \( T \to \infty \). Therefore, since \( \text{tr}(\Theta_k' H^{-1} v_i) = \text{tr}(v_i' (H^{-1})' \Theta_k' \Theta_k H^{-1}) \) and \( \Delta R_{wi,s} = \varepsilon_{i,s} + o_p(1) \), we can show that

\[ \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N} w_{1,s}^{i} \varepsilon_{i,s} (\Lambda^{-1})' H \]

\[ = \text{tr} \left( \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^{i} v_i' \varepsilon_{i,s} \frac{1}{N} \sum_{k=1}^{N} (H^{-1})' \Theta_k' \Theta_k H^{-1} \right) \]

\[ + o_p(1) = O_p \left( \frac{\sqrt{T}}{C_{NT}} \right) . \]

from which it follows that

\[ \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^{i} \varepsilon_{81,s} \Delta R_{wi,s} = O_p \left( \frac{\sqrt{T}}{C_{NT}} \right) . \]

As for the term involving \( \varepsilon_{81,s} \),

\[ \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^{i} \varepsilon_{81,s} \Delta R_{wi,s} \]

\[ \begin{aligned} &= -2 \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} d_{k,s} (\hat{\Theta}_k - \Theta_k)' \Delta x_{k,s-1} \Delta R_{wi,s} \\
&= -2 \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} \Theta_k' H^{-1} v_i (\hat{\Theta}_k - \Theta_k)' \Delta x_{k,s-1} \varepsilon_{i,s} \\
&= 2 \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} (\Delta x_{k,s-1})' \\
&= 2 \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} (\Delta x_{k,s-1})' \\
&= 2 \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} (\Delta x_{k,s-1})' \\
&= 2 \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} (\Delta x_{k,s-1})' + \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} (\Delta x_{k,s-1})' + o_p(1) \]

\[ + O_p \left( \frac{1}{\sqrt{TN^2}} \right) . \]

where the second equality uses \( \Delta R_{wi,s} = \varepsilon_{i,s} + o_p(1) \), while the third equality holds because

\[ \left| \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} d_{k,s} (\hat{\Theta}_k - \Theta_k)' \Delta x_{k,s-1} \varepsilon_{i,s} \right| \]

\[ \leq \left( \frac{1}{N} \sum_{k=1}^{N_i} ||d_k||^2 ||\hat{\Theta}_k - \Theta_k||^2 \right)^{1/2} \]

\[ \times \left( \frac{1}{N} \sum_{k=1}^{N_i} \left| \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^{i} (\Delta x_{k,s-1} \varepsilon_{i,s})^2 \right| \right)^{1/2} \]

\[ = O_p \left( \frac{1}{\sqrt{TN^2}} \right) . \]

The remaining term can be written as

\[ \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} \Theta_k' H^{-1} v_i (\hat{\Theta}_k - \Theta_k)' \Delta x_{k,s-1} \varepsilon_{i,s} \]

\[ \times \left( \Theta_k - \Theta_k \right)' H^{-1} \right) . \]

If \( k = i \), then

\[ \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} \varepsilon_{i,s} (\Delta x_{k,s-1})' \]

\[ = \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} \varepsilon_{i,s} (\Delta x_{k,s-1})' \]

\[ = O_p \left( \frac{1}{\sqrt{NT}} \right) , \]

which is \( O_p \left( \frac{1}{N} \right) \), suggesting that \( \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} \varepsilon_{i,s} (\Delta x_{k,s-1})' \sqrt{\frac{T}{NT}} \) is of the same order. If \( k \neq i \), then the same quantity is \( O_p \left( \frac{1}{\sqrt{NT}} \right) \). It follows that

\[ \left| \frac{1}{\sqrt{TN^2}} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_i} w_{1,s}^{i} \Theta_k' H^{-1} v_i (\hat{\Theta}_k - \Theta_k)' \Delta x_{k,s-1} \varepsilon_{i,s} \right| \]

\[ \leq \left( \frac{1}{N} \sum_{k=1}^{N_i} \left| \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^{i} (\Delta x_{k,s-1})' \right| \right)^{1/2} \]

\[ \times \left( \frac{1}{N} \sum_{k=1}^{N_i} ||\Theta_k||^2 ||\hat{\Theta}_k - \Theta_k||^2 \right)^{1/2} \]

\[ = O_p \left( \frac{1}{\sqrt{NT}} \right) , \]

and therefore

\[ \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^{i} \varepsilon_{81,s} \Delta R_{wi,s} = O_p \left( \frac{1}{\sqrt{NT}} \right) . \]
Thus, by combining the results,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right),
\]

the right-hand side. By cross-section independence,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

Moreover, by using the same steps as in the above, the fourth term on the right-hand side of Equation (A.19) is

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

which in turn implies

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

Consider the first term on the right-hand side of Equation (A.19) is

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

The same steps can be used to show that

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

Also,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

The order of the remainder is not the sharpest possible but it is enough to show that it is \(O_p(1/N^2)\). Consider the first term on

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]

which is clearly \(O_p(1/C_{2NT})\). As for \(\sum_{i=p_1+2}^{t} \sum_{k=1}^{N} \lambda^2_{i,k} \mathcal{C}_{1,k} \Delta R_{wi,k} \sqrt{T} \), since \(\epsilon_{i,s} \) is independent across \(i\) and also of \(F_i, \sum_{i=p_1+2}^{t} \sum_{k=1}^{N} \lambda^2_{i,k} \mathcal{C}_{1,k} \Delta R_{wi,k} \sqrt{T} \),

\[
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} \left( \lambda^2_{1,s} \mathcal{C}_{1,s} \Delta R_{wi,s} \right) = O_p \left( \frac{1}{\sqrt{T}} \right).
\]
Thus, by adding the terms, \( F_i F_i' w^3_{i,s} e_{i,s}^2 / \sqrt{T} = O_p(1) \). Hence, since \( \text{tr}(d_k F_i F_i') = \text{tr}(d_k d_k' F_i F_i') \), we get

\[
\frac{1}{\sqrt{T} N_1} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w^3_{i,s} e_{i,s}^2 (d_k' H F_s) = \text{tr} \left( \frac{1}{\sqrt{T} N_1} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} d_k d_k' F_i F_i' w^3_{i,s} e_{i,s}^2 \right)
\]

\[
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||d_k||^2 \right)^{1/2} \times \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left[ \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t F_i F_i' w^3_{i,s} e_{i,s}^2 \right]^2 \right)^{1/2}
\]

\[
= O_p \left( \frac{1}{C_{2NT}} \right),
\]

and therefore

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w^5_{i,s} c_{i,s} e_{i,s} = O_p \left( \frac{1}{C_{2NT}} \right).
\]

Finally, consider \( \sum_{s=p_i+2}^t w^5_{i,s} c_{i,s} e_{i,s} / \sqrt{T} \). Again, since \( e_{i,s} \) is independent across \( s \) and also of \( F_i, \), \( \sum_{s=p_i+2}^t F_i \Delta x_{i,s-1} w^3_{i,s} e_{i,s} / \sqrt{T} = O_p(1) \), from which we obtain

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w^5_{i,s} c_{i,s} e_{i,s} = O_p \left( \frac{1}{\sqrt{T} C_{2NT}} \right).
\]

Thus, by adding the terms,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w^5_{i,s} c_{i,s} \Delta R_{i,s} = O_p \left( \frac{1}{C_{2NT}} \right).
\]

The sixth and last term on the right-hand side of Equation (A.19), \( \sum_{s=p_i+2}^t w^5_{i,s} c_{i,s}(a_{i,s} - (\Phi_i - \Phi_i') \Delta x_{i,s-1}) / \sqrt{T} \) is dominated by the other remainders and can be treated as \( o_p(1/C_{2NT}) \). Therefore, by adding Equations (A.19)–(A.24),

\[
\frac{1}{\sqrt{T}} \hat{R}_{i,t} = \frac{1}{\sqrt{T}} R_{i,t} + \frac{1}{\sqrt{T}} G_{i,t} + O_p \left( \frac{1}{C_{2NT}} \right),
\]

where

\[
G_{i,t} = \Theta_i' H^{-1} \sum_{s=p_i+2}^t w_{i,s} v_s - \frac{1}{2} \sum_{s=p_i+2}^t w^3_{i,s} \bar{c}_{01,s} e_{i,s}.
\]

This establishes the second result, and hence the proof of the lemma is complete.

**Proof of Theorem 1.** Consider again the group consisting of the first \( N_1 \) observations. By Lemma A.1, with \( k = 0 \),

\[
\Delta \hat{R}_{i,t} = \Delta R_{i,t} + g_{i,t} + O_p \left( \frac{1}{C_{2NT}} \right)
\]

\[
= \Delta R_{i,t} + O_p \left( \frac{1}{C_{NT}} \right).
\]

As for \( \hat{R}_{i,t} / \sqrt{T} \), we have, by a simple modification of Lemma 2 in Bai and Ng (2004),

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s} v_s = O_p \left( \frac{1}{\sqrt{N}} \right) + O_p \left( \frac{1}{T^{3/4}} \right).
\]

Hence, since by the proof of Lemma A.1, \( \sum_{s=p_i+2}^t w^3_{m,s} c_{0m,s} e_{i,s} / \sqrt{T} = O_p(1/\sqrt{N}) \),

\[
\frac{1}{\sqrt{T}} \hat{R}_{i,t} = \frac{1}{\sqrt{T}} (R_{i,t} + G_{i,t}) + O_p \left( \frac{1}{C_{2NT}} \right)
\]

\[
= \frac{1}{\sqrt{T}} R_{i,t} + O_p \left( \frac{1}{C_{NT}} \right).
\]

Consider \( R_{i,t} / \sqrt{T} \). Since \( R_{i,t} = \rho_i R_{i,t-1} + e_{i,t} \), where \( \rho_i = \exp(c_i/T) \) and \( e_{i,t} = \Theta_i' F_i + e_{i,t} \), by recursive substitution,

\[
\frac{1}{\sqrt{T}} R_{i,t} = \frac{1}{\sqrt{T}} \sum_{k=p_i+2}^t \exp \left( \frac{c_i}{T} \right)^{-k} e_{i,k} + o_p(1)
\]

\[
= \Theta_i' \frac{1}{\sqrt{T}} \sum_{k=p_i+2}^t \exp \left( \frac{c_i}{T} \right)^{-k} F_k + \frac{1}{\sqrt{T}}
\]

\[
\times \sum_{k=p_i+2}^t \exp \left( \frac{c_i}{T} \right)^{-k} e_{i,k} + o_p(1),
\]

from which we obtain

\[
\frac{1}{\sqrt{T}} R_{i,t} \to_d \Theta_i' B_{F,i}(s) + B_{e,i}(s) = B_{R,i}(s)
\]

as \( T \to \infty \), where

\[
B_{e,i}(r) = \int_{s=0}^r \exp((r-s)c_i) \sigma_i(s) dW_{e,i}(s),
\]

\[
B_{F,i}(r) = \int_{s=0}^r \exp((r-s)c_i) \Sigma(s)^{1/2} dW_{F}(s),
\]
with $W_{t,i}(s)$ and $W_{t}(s)$ being two independent Brownian motions. Hence, since $R_{wi,t} = \sum_{t'=p_i+2}^{T} \Delta R_{wi,k}$, we obtain

$$
\frac{1}{\sqrt{T}} \hat{R}_{wi,t} = \frac{1}{\sqrt{T}} \sum_{k=p_i+2}^{T} E_{i,k} + c_i \frac{T^{3/2}}{3}
\times \sum_{k=p_i+2}^{T} w_{1,k} R_{i,k-1} + O_p \left( \frac{1}{C_{NT}} \right)
\rightarrow_d W_{t,i}(s) + c_i \int_{t=0}^{s} w_{1}(r) B_{R_i}(r) dr = V_{R_i}(s)
$$

(A.30)

as $N, T \rightarrow \infty$. This result, together with the continuous mapping theorem, imply

$$
\frac{1}{T} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t} = \frac{1}{T} \sum_{t=p_i+2}^{T} R_{wi,t-1} \Delta R_{wi,t} + O_p \left( \frac{1}{C_{NT}} \right)
\rightarrow_d \int_{s=0}^{1} V_{R_i}(s)^2 ds
$$

(A.31)

The proof is completed by noting that

as $N, T \rightarrow \infty$. □

**Proof of Theorem 2.** Consider the numerator of $\tau$.

By definition, $\hat{R}_{wi,t} = \sum_{t'=p_i+2}^{T} \Delta \hat{R}_{wi,s}$ or $\hat{R}_{wi,t} = \hat{R}_{wi,t-1} + \Delta \hat{R}_{wi,t}$, showing that $\hat{R}_{wi,t}^2 = (\hat{R}_{wi,t-1} + \Delta \hat{R}_{wi,t})^2 = \hat{R}_{wi,t-1}^2 + (\Delta \hat{R}_{wi,t})^2$. It follows that

$$
\sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t} = \frac{1}{2} \sum_{t=p_i+2}^{T} (\hat{R}_{wi,t}^2 - \hat{R}_{wi,t-1}^2) - \frac{1}{2} \sum_{t=p_i+2}^{T} (\Delta \hat{R}_{wi,t})^2
\leq \frac{1}{2} (\hat{R}_{wi,t} - \hat{R}_{wi,p_i+1}) - \frac{1}{2} \sum_{t=p_i+2}^{T} (\Delta \hat{R}_{wi,t})^2,
$$

(A.33)

with a similar result applying to $\sum_{t=p_i+2}^{T} R_{wi,t-1} \Delta R_{wi,t}$. Write

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t}
= \frac{1}{\sqrt{NT}} \sum_{m=1}^{M} \sum_{i=N_{m-1}+1}^{N_m} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t}
\leq \left( \frac{1}{T} \sum_{t=p_i+2}^{T} \sum_{i=1}^{N_i} \left( \frac{1}{\sqrt{NT}} \right) \right)^{1/2} \left( \frac{1}{T} \sum_{t=p_i+2}^{T} \sum_{i=1}^{N_i} \frac{R_{wi,t}}{\Theta_{i}} \right)^{1/2}
\leq O_p \left( \frac{1}{\sqrt{N}} \right).
$$

Hence, focusing of the first group comprising the first $N_1$ units,

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t}
= \frac{1}{2 \sqrt{NT}} \sum_{i=1}^{N_i} (\hat{R}_{wi,T}^2 - \hat{R}_{wi,p_i+1}^2)
\leq \frac{1}{2 \sqrt{NT}} \sum_{i=1}^{N_i} (\Delta \hat{R}_{wi,t})^2.
$$

(A.34)

It is trivial to show that $\sum_{i=1}^{N_i} (\hat{R}_{wi,T}^2 - \hat{R}_{wi,p_i+1}^2) \sqrt{NT}$ and $\sum_{i=1}^{N_i} (\Delta \hat{R}_{wi,t})^2 \sqrt{NT}$ are $O_p(\sqrt{N/T})$. Let us therefore consider $\hat{R}_{wi,T}^2$. Clearly, by Lemma A.1,

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \hat{R}_{wi,T}^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} (R_{wi,T}^2 + 2 R_{wi,T} G_{i,T} + G_{i,T}^2)
= O_p \left( \frac{\sqrt{N}}{C_{2NT}} \right),
$$

where

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} R_{wi,T} G_{i,T}
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{s=p_i+2}^{T} w_{s,1} v_s
\leq \frac{1}{2 \sqrt{NT}} \sum_{i=1}^{N_i} \sum_{s=p_i+2}^{T} w_{s,1}^2 c_{01,s} e_{i,s}
\leq O_p \left( \frac{1}{T^{3/2}} \right).
$$

As for the first term on the right-hand side,

$$
\frac{1}{\sqrt{T}} \sum_{i=1}^{N_i} R_{wi,T} \Theta_i
= \frac{1}{\sqrt{T}} \sum_{i=1}^{N_i} \sum_{s=p_i+2}^{T} \epsilon_{i,s} \Theta_i + \frac{1}{T^{3/2} N}
\times \sum_{i=1}^{N_i} \sum_{s=p_i+2}^{T} c_i \Theta_i \epsilon_{i,s} R_{i,s-1}
\leq O_p(1).
$$

Moreover, as in the proof of Theorem 1,

$$
\sum_{s=p_i+2}^{T} w_{s,1} v_s \sqrt{T} = O_p(\sqrt{N}) + O_p(1/T^{3/2}) = O_p(1/C_{NT}),
$$

and since $c_{01,s} = O_p(\sqrt{N})$, we also have

$$
\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} w_{t,1}^2 c_{01,s} e_{i,s} \right|
\leq \left( \frac{1}{T} \sum_{s=p_i+2}^{T} \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \left( \frac{1}{\sqrt{NT}} \right) \right) \right)^{1/2} \left( \frac{1}{T} \sum_{s=p_i+2}^{T} \sum_{i=1}^{N_i} w_{s,1}^2 c_{01,s} \right)^{1/2}
= O_p \left( \frac{1}{\sqrt{N}} \right).
$$
It follows that
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} R_{wi,t} G_{i,t} = O_p \left( \frac{1}{C_{NT}} \right). \]

Also, since by the proof of Lemma A.1, \( \sum_{i=p_i+2}^{T} w_{i,t}^3 \bar{c}_{i} = O_p(1/\sqrt{N}) \), suggesting
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} G_{i,t}^2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N_t} \left( \Theta_i^{-1} H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{T} w_{s,t} v_s \right) \]
\[ - \frac{1}{2\sqrt{T}} \sum_{s=p_i+2}^{T} w_{s,t}^3 \bar{c}_{i} \epsilon_{i,s}^2 = O_p \left( \frac{\sqrt{N}}{C_{NT}} \right). \]

It follows that
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \hat{R}_{wi,t}^2 = O_p \left( \frac{\sqrt{N}}{C_{2NT}} \right). \quad (A.35) \]

Next, consider \( \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} (\hat{R}_{wi,t})^2 / \sqrt{NT} \) in Equation (A.33). By Lemma A.1,
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} (\hat{R}_{wi,t})^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} \left( (\hat{R}_{wi,t})^2 - 2\hat{R}_{wi,t} g_{i,t} + g_{i,t}^2 \right) \]
\[ + O_p \left( \frac{\sqrt{N}}{C_{2NT}} \right), \]

where
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} \Delta R_{wi,t} g_{i,t} \]
\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} \Delta R_{wi,t} \Theta_i^{-1} H^{-1} w_{i,t} v_i \]
\[ - \frac{1}{2} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} w_{i,t}^3 (\bar{c}_{0i,t} + \bar{c}_{4i,t} + \bar{c}_{6i,t}) \epsilon_{i,t}. \]

From Lemma A.1 we have that \( \sum_{i=1}^{N_t} \Delta R_{wi,t} \Theta_i^{-1} / \sqrt{N} = O_p(1) \) and \( ||v_i|| = O_p(1/C_{1NT}) \), and therefore
\[ \left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} \Delta R_{wi,t} \Theta_i^{-1} H^{-1} w_{i,t} v_i \right| \]
\[ \leq \left( \frac{1}{T} \sum_{s=p_i+2}^{T} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N_t} \Delta R_{wi,t} \Theta_i^{-1} \right|^2 \right)^{1/2} \]
\[ \times \left( \frac{1}{T} \sum_{s=p_i+2}^{T} ||w_{1,t}||^2 ||v_i||^2 \right)^{1/2} \]
\[ = O_p \left( \frac{1}{C_{1NT}} \right). \]

Similarly, since \( (\bar{c}_{01,t} + \bar{c}_{41,t} + \bar{c}_{61,t}) = O_p(1/C_{NT}) \) and
\[ \sum_{i=1}^{N_t} \epsilon_{i,t} / \sqrt{N} = O_p(1), \] we can show that
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} w_{i,t} (\bar{c}_{01,t} + \bar{c}_{41,t} + \bar{c}_{61,t}) \epsilon_{i,t} = O_p \left( \frac{1}{C_{NT}} \right). \]

From which we obtain
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} (\hat{R}_{wi,t}^2) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} (\hat{R}_{wi,t})^2 + O_p \left( \frac{\sqrt{N}}{C_{2NT}} \right). \quad (A.36) \]

Thus, by adding the results,
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t} + O_p \left( \frac{\sqrt{N}}{C_{2NT}} \right). \quad (A.37) \]

Consider \( \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} R_{wi,t-1} \Delta \hat{R}_{wi,t} / \sqrt{NT} \). By using \( \kappa = 1/2 \), and the definition of \( \Delta \hat{R}_{wi,t} \),
\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} R_{wi,t-1} \Delta \hat{R}_{wi,t} \]
\[ = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \]
\[ + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_t} \sum_{t=p_i+2}^{T} R_{wi,t-1} \epsilon_{i,t}. \quad (A.38) \]

As for the first term, since \( \kappa = 1/2 \) in this case, \( \rho_t = \exp(c_i / \sqrt{NT}) = 1 + c_i / \sqrt{NT} + o_p(1) \), which in turn suggests
\[ E \left( \frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} R_{wi,t-1} R_{i,t-1} | c_i \right) \]
\[ = \frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} \sum_{s=p_i+2}^{T} E(\epsilon_{i,s} R_{i,s-1} | c_i) + \frac{1}{\sqrt{NT}} \]
\[ \times \sum_{t=p_i+2}^{T} c_i w_{1,t} \sum_{s=p_i+2}^{T} w_{1,s} E(R_{i,s-1} R_{i,s-1} | c_i) + o_p(1). \quad (A.39) \]

Consider \( R_{i,t} / \sqrt{T} \). By substitution of \( \rho_t = 1 + c_i / \sqrt{NT} + o_p(1) \),
\[ \frac{1}{\sqrt{T}} R_{i,t} = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{T} \rho_t^{-s} \epsilon_{i,t} + o_p(1) \]
\[ = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{T} \epsilon_{i,t} + o_p(1). \]
Moreover, since $\varepsilon_{i,t}$ and $F_i$ are independent, $E(\varepsilon_{i,s} \varepsilon_{i,k}) = E(\varepsilon_{i,s} \Theta_i F_i + \varepsilon_{i,k}) = \sigma_{1,k} E(\varepsilon_{i,s} \varepsilon_{i,k})$, which is zero if $s \neq k$ and $\sigma_{1,k}$ if $s = k$. Letting

$$g_1(v) = w_1(v) \int_{u=0}^{v} \sigma_1(r)du,$$

these results imply

$$\frac{1}{T^2} \sum_{t=p_0+2}^{T} \sum_{s=p_0+2}^{t-1} w_{1,t} w_{1,s} E(\varepsilon_{i,s} \varepsilon_{i,t-1}) = \frac{1}{T^2} \sum_{t=p_0+2}^{T} \sum_{s=p_0+2}^{t-1} \sum_{k=p_0+2}^{t-1} E(\varepsilon_{i,s} \varepsilon_{i,k}) = \frac{1}{T^2} \sum_{t=p_0+2}^{T} \sum_{s=p_0+2}^{t-1} \sum_{k=p_0+2}^{t-1} \sigma_{1,k} \to \int_{v=0}^{1} g_1(v)dv$$

as $T \to \infty$. But we also have $E(\varepsilon_{i,s} \varepsilon_{i,k}) = \Theta_i E(F_i F_i') + \sigma_{1,k} \sigma_{1,k}$, which is zero if $s \neq k$ and $(\Theta_i' \Theta_i + \sigma_{1,k}^2)$ if $s = k$. Hence, with $t \geq s$,

$$E(R_{i,s} R_{i,t}) = \sum_{k=p_0+2}^{t} \sum_{n=p_0+2}^{t} E(\varepsilon_{i,k} \varepsilon_{i,n}) = \sum_{k=p_0+2}^{t} E(\varepsilon_{i,k}^2) = \sum_{k=p_0+2}^{t} (\Theta_i' \Sigma_k \Theta_i + \sigma_{1,k}^2),$$

which we can use to show that

$$\frac{1}{T^2} \sum_{t=p_0+2}^{T} \sum_{s=p_0+2}^{t-1} w_{1,t} w_{1,s} E(R_{i,s-1} R_{i,t-1} | c_i) = \frac{1}{T^2} \sum_{t=p_0+2}^{T} \sum_{s=p_0+2}^{t-1} \sum_{k=p_0+2}^{t-1} (\Theta_i' \Sigma_k \Theta_i + \sigma_{1,k}^2) = \int_{u=0}^{1} g_1(u,v)dvdu$$

as $T \to \infty$, where

$$g_1(u,v) = w_1(u) w_1(v) \int_{x=0}^{v} (\Theta_i' \Sigma(x) \Theta_i + \sigma_1(x)^2)dx,$$

Let us also define

$$q_{N,i}(u) = \mu \gamma_1(u) + \frac{(\mu_2^2 + \sigma_2^2)}{\sqrt{N}} \int_{v=0}^{u} g_i(u,v)dv,$$

$$\bar{q}_{N,i}(u) = \frac{1}{N_i} \sum_{i=1}^{N_i} q_{N,i}(u),$$

with $q_i(u)$ and $\bar{q}_i(u)$ being the corresponding integrals after taking $N \to \infty$. That is, $q_i(u) = \bar{q}_i(u) = \mu \gamma_1(u)$. By using this and the fact that $E(c_i^2) = \mu_2^2 + \sigma_2^2$, we obtain

$$E\left(\frac{1}{T^2} \sum_{t=p_0+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \right) \to \int_{u=0}^{1} q_i(u)du,$$

and by application of Corollary 1 of Phillips and Moon (1999), with $[\lambda_i N] = N_1$, we can further show that

$$\frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{t=p_0+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \to p \lambda_i \int_{u=0}^{1} \bar{q}_1(u)du$$

as $N, T \to \infty$, where $p$ signifies convergence in probability. Thus, considering the whole cross-section,

$$\frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{t=p_0+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \to \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \frac{1}{(N_m - N_{m-1})T^2} \sum_{i=N_{m-1}+1}^{N_m} \sum_{t=p_0+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \to p \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_{u=0}^{1} \bar{q}_m(u)du.$$ (A.42)

Consider $\sum_{i=1}^{N_1} \sum_{t=p_0+2}^{T} R_{wi,t-1} \varepsilon_{i,t}/\sqrt{NT}$. Clearly, $E(R_{wi,t-1} \varepsilon_{i,t}) = 0$. As for the variance,

$$E\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_0+2}^{T} R_{wi,t-1} \varepsilon_{i,t}\right)^2$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{t=p_0+2}^{T} \sum_{j=p_0+2}^{T} E(R_{wi,t-1} R_{wi,j-1}) E(\varepsilon_{i,t} \varepsilon_{j,t})$$

$$= \frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{t=p_0+2}^{T} E(R_{wi,t-1}^2).$$ (A.43)

Consider $E(R_{wi,t-1}^2)$. Because of symmetry,

$$\sum_{s=p_0+2}^{t-1} \sum_{k=p_0+2}^{s-1} w_{1,s} w_{1,k} E(R_{i,s-1} R_{i,k-1} | c_i)$$

$$= \sum_{s=p_0+2}^{t-1} w_{1,s} E(R_{i,s-1}^2 | c_i)$$

$$+ 2 \sum_{s=p_0+2}^{t-1} \sum_{k=p_0+2}^{s-1} w_{1,s} w_{1,k} E(R_{i,s-1} R_{i,k-1} | c_i)$$

$$= \sum_{s=p_0+2}^{t-1} w_{1,s} E(R_{i,s-1}^2 | c_i)$$

$$+ 2 \sum_{s=p_0+2}^{t-1} \sum_{k=p_0+2}^{s-1} w_{1,s} w_{1,k} \sum_{n=p_0+2}^{s} (\Theta_i' \Sigma_k \Theta_i + \sigma_{1,n}^2)$$

$$\to \int_{u=0}^{1} \gamma_2(u,v)dv + 2 \int_{u=0}^{1} \gamma_1(u,v)dv du.$$
as $T \to \infty$. Therefore,

$$
\frac{1}{T} E \left( R_{wi,t-1}^2 | c_i \right) = 
\frac{1}{T} \sum_{s=p_i+2}^{T} \sum_{k=p_i+2}^{T} E(\varepsilon_{i,s}\varepsilon_{i,k}) + \frac{2c_i^2}{\sqrt{NT}} \sum_{s=p_i+2}^{T} \sum_{k=p_i+2}^{T} w_{1,s} E(R_{i,s-1} \varepsilon_{i,k} | c_i) + \frac{c_i^2}{NT^2} \sum_{s=p_i+2}^{T} \sum_{k=p_i+2}^{T} w_{1,s} w_{1,k} E(R_{i,s-1} R_{i,k-1} | c_i)

\to r + \frac{2c_i}{\sqrt{N}} \int_0^r g_i(u) du + \frac{c_i^2}{N} \int_0^r \left( g_i(u, u) + 2 \int_0^u g_i(u, v) dv \right) du,
$$

suggesting that

$$
\frac{1}{T} E \left( R_{wi,t-1}^2 \right) = 
\frac{1}{T} \sum_{s=p_i+2}^{T} \sum_{k=p_i+2}^{T} E(R_{wi,t-1}^2 | c_i) \to r + \frac{2\mu_c}{\sqrt{N}} \int_0^r g_i(u) du + \frac{\mu_c^2 + \sigma_i^2}{N} \int_0^r \left( g_i(u, u) + 2 \int_0^u g_i(u, v) dv \right) du = p_{N,i}(r),
$$

with an obvious definition of $p_{N,i}(r)$. The limit of (A.43) is therefore given by

$$
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} R_{wi,t-1} \varepsilon_{i,t} \right)^2 \right] = \frac{1}{NT^2} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} E \left( R_{wi,t-1}^2 \right) \to \lambda_1 \int_0^1 p_{N,i}(r) dr
$$

as $T \to \infty$.

The conditions of Theorem 2 of Phillips and Moon (1999) hold. Hence, using $p_i(r)$ to denote the limit of $p_{N,i}(r)$,

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} R_{wi,t-1} \varepsilon_{i,t} \to_d N \left( 0, \lambda_1 \int_0^1 p_i(r) dr \right)
$$

as $N, T \to \infty$. Note also that since the effect of the common component in $p_{N,i}(r)$ vanishes as $N \to \infty$, $R_{wi,t-1} \varepsilon_{i,t}$ is asymptotically uncorrelated across groups, and thus independent by normality. The same is true for $\hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t}$ if we assume that $N/T \to 0$ such that $O_p(\sqrt{N/C_{2NT}}) = o_p(1)$. It follows that

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t} \to_d N \left( 0, \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_0^1 \overline{p}(r) dr \right).
$$

as $N, T \to \infty$ with $N/T \to 0$ and where the definition of $\overline{p}_i(u)$ is analogous to that of $\overline{p}_i(u)$.

As for the denominator of $\tau$, by using the same steps as before, via Corollary 1 of Phillips and Moon (1999),

$$
\frac{1}{NT^2} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1}^2 = \frac{1}{NT^2} \sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} R_{wi,t-1}^2 + O_p \left( \frac{1}{C_{2NT}} \right)

\to \frac{1}{N} \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_0^1 \overline{p}_m(r) dr
$$

as $N, T \to \infty$. Hence, letting

$$
\theta = \frac{\sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_0^1 \overline{q}_{N,m}(r) dr}{\sqrt{\sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_0^1 \overline{p}_m(r) dr}}

= \lim_{N \to \infty} \frac{\sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_0^1 \overline{q}_{N,m}(r) dr}{\sqrt{\sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_0^1 \overline{p}_m(r) dr}}
$$

by the continuous mapping theorem,

$$
\tau = \frac{\theta}{\sqrt{\sum_{i=1}^{N_i} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t-1}^2}} \int_0^1 \Delta \hat{R}_{wi,t}^2
\to_d \left( \theta + N(0, 1) \right)^2 =_d \chi_1^2(\theta^2)
$$

as $N, T \to \infty$ with $N/T \to 0$, where $=_d$ signifies equality in distribution.

**Proof of Proposition 1.** For notational simplicity, in this proof, when context is clear, the dependence of $\hat{\sigma}_{1,i}(n)$ and $\hat{\sigma}_{2,i}(n)$ on $n$ will be suppressed.

The criterion function is

$$
QML(n) = \sum_{t=p_i+2}^{T} [n \log(\hat{\sigma}_{1,i}(n)) + (N-n) \log(\hat{\sigma}_{2,i}(n))],
$$

Define

$$
QML_0(n) = \sum_{t=p_i+2}^{T} [n \log(\sigma_{1,i}) + (N-n) \log(\sigma_{2,i})].
$$
such that
\[
\text{QML}(n) - \text{QML}_0(N_1) = \sum_{t=p_1+2}^T \left[ (n \log (\hat{\sigma}_{1,t}^2(n)) - N_1 \log (\sigma_{1,t}^2) ) + ((N - n) \times \log (\hat{\sigma}_{2,t}^2(n)) - (N - N_1) \log (\sigma_{2,t}^2) ) \right]
\]
\[
= \sum_{t=p_1+2}^T \left[ n \log \left( \frac{\hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} \right) + (N - n) \log \left( \frac{\hat{\sigma}_{2,t}^2(n)}{\sigma_{2,t}^2} \right) \right]
\]
\[
+ \left( (N - n) \log \left( \frac{\hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} \right) - (N - N_1) \log (\sigma_{2,t}^2) \right] \]
\[
= \sum_{t=p_1+2}^T \left[ n \log \left( \frac{\hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} \right) + (N - n) \log \left( \frac{\hat{\sigma}_{2,t}^2(n)}{\sigma_{2,t}^2} \right) \right]
\]
\[
+ (n - N_1) \log \left( \frac{\sigma_{1,t}^2}{\sigma_{2,t}^2} \right) \right].
\]  
(A.49)

As in the proof of Proposition 1, we start with
\[
\hat{\sigma}_{m,t}^2 = \sigma_{m,t}^2 + \tau_{0m,t} + O_P \left( \frac{1}{C_{2NT}} \right) = \sigma_{m,t}^2 + O_P \left( \frac{1}{\sqrt{N}} \right),
\]
where \( C_{2NT} = \min(N, \sqrt{T}) \) and \( \tau_{0m,t} = \sum_{i=N_{m-1}+1}^{N_m} \left( x_{i,t} - \sigma_{m,t}^2 \right) / \left( N_m - N_{m-1} \right) \). This, together with Taylor expansion of the type \( \log(1 + x) = x + O(x^2) \), gives
\[
\log \left( \frac{\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2} \right) = \log \left( 1 + \frac{\hat{\sigma}_{m,t}^2 - \sigma_{m,t}^2}{\sigma_{m,t}^2} \right)
\]
\[
= \frac{\hat{\sigma}_{m,t}^2 - \sigma_{m,t}^2}{\sigma_{m,t}^2} + O_P \left( \left( \frac{\hat{\sigma}_{m,t}^2 - \sigma_{m,t}^2}{\sigma_{m,t}^2} \right)^2 \right)
\]
\[
= \frac{\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2} - 1 + O_P \left( \tau_{0m,t}^2 \right),
\]
where \( \hat{\sigma}_{m,t}^2 / \sigma_{m,t}^2 - 1 = \tau_{0m,t}^2 / \sigma_{m,t}^2 = O_P \left( 1 / \sqrt{N} \right) \) and \( O_P (\tau_{0m,t}^2) = O_P \left( 1 / N \right) \). Hence, since \( \hat{\sigma}_{m,t}^2 / \sigma_{m,t}^2 - 1 \) is dominated by \( O_P (\tau_{0m,t}^2) \), we can show that
\[
\text{QML}(n) - \text{QML}_0(N_1) \sim \sum_{t=p_1+2}^T \left[ \frac{n \hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} + (N - n) \frac{\hat{\sigma}_{2,t}^2(n)}{\sigma_{2,t}^2} - N \right]
\]
\[
+ (n - N_1) \log \left( \frac{\sigma_{1,t}^2}{\sigma_{2,t}^2} \right). \]  
(A.50)

We now show that the objective function cannot achieve its minimum unless \( n = N_1 \). Because of symmetry, we only need to consider the case when \( n \leq N_1 \). Again, from the proof of Lemma A.1, replacing \( N_1 \) by \( n \),
\[
\hat{\sigma}_{1,t}^2(n) = \frac{1}{n} \sum_{i=1}^n \epsilon_{i,t}^2 + O_P \left( \frac{1}{C_{2NT}} \right)
\]
\[
= \frac{1}{n} \sum_{i=1}^n \epsilon_{i,t}^2 + O_P \left( \frac{1}{C_{2NT}} \right),
\]
and
\[
\hat{\sigma}_{2,t}^2(n) = \frac{1}{(N - n)} \sum_{i=n+1}^N \epsilon_{i,t}^2 + O_P \left( \frac{1}{C_{2NT}} \right)
\]
\[
= \frac{1}{(N - n)} \sum_{i=n+1}^N \epsilon_{i,t}^2 + \sum_{i=N_1+1}^N \epsilon_{i,t}^2 + O_P \left( \frac{1}{C_{2NT}} \right),
\]
from which we obtain
\[
\text{QML}(n) - \text{QML}_0(N_1) \sim \sum_{t=p_1+2}^T \left[ \sum_{i=1}^n \epsilon_{i,t}^2 + \sum_{i=n+1}^N \epsilon_{i,t}^2 - N \right]
\]
\[
+ (n - N_1) \log \left( \frac{\sigma_{1,t}^2}{\sigma_{2,t}^2} \right) \]  
(A.51)

\[
= \sum_{t=p_1+2}^T \left[ \sum_{i=1}^n \epsilon_{i,t}^2 + \sum_{i=n+1}^N \epsilon_{i,t}^2 - N \right] + (n - N_1) \log \left( \frac{\sigma_{1,t}^2}{\sigma_{2,t}^2} \right)
\]
\[
+ \frac{1}{\sqrt{NT}} \sum_{i=n+1}^N \sum_{t=p_1+2}^T x_t \epsilon_{i,t}^2 - 1 \]  
(A.52)

From here it is easy to see that if \( n = N_1 \), then
\[
\frac{1}{\sqrt{NT}} \text{QML}(n) - \text{QML}_0(N_1) \sim \frac{1}{\sqrt{NT}} \sum_{t=p_1+2}^T \sum_{i=1}^N \epsilon_{i,t}^2 - 1 = O_p(1),
\]
whereas if \( n < N_1 \), then
\[
\frac{1}{\sqrt{NT}}[\text{QML}(n) - \text{QML}_0(N_1)]
\sim \sqrt{NT} \frac{(N_1 - n)}{NT} \sum_{t=1}^{T} [x_t - \log(1 + x_t)] + O_p(1)
\]
\[= O_p(\sqrt{NT}). \] (A.53)

Note that QML_0(N_1) is independent of \( n \). Hence, minimizing QML(n) is equivalent to minimizing \([\text{QML}(n) - \text{QML}_0(N_1)]\). Because of this, and the fact that the order of magnitude of \([\text{QML}(n) - \text{QML}_0(N_1)]\) increases by a factor of \(\sqrt{NT}\) when \(n < N_1\), the minimum is obtained by setting \(n = N_1\), as was to be shown.

[Received December 2012. Revised September 2013]

REFERENCES

Abadir, K. M., and Distaso, W. (2007), “Testing Joint Hypotheses When One of the Alternatives is One-Sided,” *Journal of Econometrics*, 140, 695–718. [115]

Ahn, S. K. (1993), “Some Tests for Unit Roots in Autoregressive-Integrated-Moving Average Models With Deterministic Trends,” *Biometrika*, 80, 855–868. [114,115]

Bai, J. (1997), “Estimating Multiple Breaks One at a Time,” *Econometric Theory*, 13, 315–352. [117,118,119]

——— (2003), “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71, 135–173. [118,122,126]

Bai, J., and Ng, S. (2002), “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221. [118,119]

——— (2004), “A Panic Attack on Unit Roots and Cointegration,” *Econometrica*, 72, 1127–1177. [114,119,122,126,129]

Bai, J., and Perron, P. (1998), “Estimating and Testing Linear Models With Multiple Structural Change,” *Econometrica*, 66, 47–78. [117]

Beare, B. K. (2008), “Unit Root Testing With Unstable Volatility,” Nuffield College Economics Working Paper No. 2008-06, Oxford University. [112,114,115]

Boswijk, H. P. (2005), “Adaptive Testing for a Unit Root With Nonstationary Volatility,” *UvA-Econometrics Discussion Paper 2005/07*. [116]

Busetti, F., and Taylor, A. M. R. (2003), “Variance Shifts, Structural Breaks and Stationarity Tests,” *Journal of Business & Economic Statistics*, 21, 510–531. [112]

Cavaliere, G., Phillips, P. C. B., Smeekes, S., and Taylor, A. M. R. (2012), “Lag Length Selection for Unit Root Tests in the Presence of Nonstationary Volatility,” *Cowles Foundation Discussion Article No. 1844*. [118]

Cavaliere, G., and Taylor, A. M. R. (2009a), “Testing for a Change in Persistence in the Presence of Non-Stationary Volatility,” *Journal of Econometrics*, 147, 84–98. [112]

——— (2009b), “Heterostastic Time Series With a Unit Root,” *Econometric Theory*, 25, 1228–1276. [112,116,118]

Chauvet, M., and Potter, S. M. (2001), “Recent Changes in the U.S. Business Cycle,” *The Manchester School*, 69, 481–508. [112]

Demetrescu, M., and Hanck, C. (2012), “Unit Root Testing in Heterostastic Panels Using the Cauchy Estimator,” *Journal of Business and Economic Statistics*, 30, 256–264. [112,116,117,118,121]

Kim, C.-J., and Nelson, C. R. (1999), “Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle,” *Review of Economics and Statistics*, 81, 608–616. [112]

Koop, G., and Potter, S. M. (2000), “Nonlinearity, Structural Breaks or Outliers in Economic Time Series?,” in *Nonlinear Econometric Modelling in Time Series Analysis*, eds. W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjostheim, and A. H. Wurtz, Cambridge: Cambridge University Press, pp. 61–78. [112]

Lin, C.-C., and Ng, S. (2012), “Estimation of Panel Data Models With Parameter Heterogeneity When Group Membership Is Unknown,” *Journal of Econometric Methods*, 1, 42–55. [118]

Lopez, J. A. (2001), “Evaluating the Predictive Accuracy of Volatility Models,” *Journal of Forecasting*, 20, 87–109. [115]

Loretan, M., and Phillips, P. C. B. (1994), “Testing Covariance Stationarity Under Moment Condition Failure With an Application to Common Stock Returns,” *Journal of Empirical Finance*, 1, 211–248. [112]

McConnell, M. M., and Perez Quiros, G. (2000), “Output Fluctuations in the United States: What Has Changed Since the Early 1980s?”, *American Economic Review*, 90, 1464–1476. [112]

McMillan, D. G., and Ruiz, I. (2009), “Volatility Persistence, Long Memory and Time-Varying Unconditional Mean: Evidence From 10 Equity Indices,” *The Quarterly Review of Economics and Finance*, 49, 578–595. [112]

Moon, H. R., Perron, B., and Phillips, P. C. B. (2007), “Incidental Trends and the Power of Panel Unit Root Tests,” *Journal of Econometrics*, 141, 416–459. [117]

Ng, S., and Perron, P. (1995), “Unit Root Test in ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag.” *Journal of the American Statistical Association*, 90, 268–281. [119]

——— (2001), “Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power,” *Econometrica*, 69, 1519–1554. [118]

Pagan, A. R., and Schwert, G. W. (1990), “Testing for Covariance Stationarity in Stock Market Data,” *Economics Letters*, 33, 165–170. [112]

Perron, P. (2006), “Dealing With Structural Breaks,” in *Palgrave Handbook of Econometrics, volume 1: Econometric Theory*, eds. K. Patterson and T. C. Mills, Basingstoke, UK: Palgrave Macmillan, pp. 278–352. [118]

Phillips, P. C. B., and Moon, R. H. (1999), “Linear Regression Limit Theory of Nonstationary Panel Data,” *Econometrica*, 67, 1057–1111. [152,153]

Pesaran, H. M. (2007), “A Simple Panel Unit Root Test in the Presence of Cross Section Dependence,” *Journal of Applied Econometrics*, 22, 265–312. [114,116]

Pesaran, H. M., Smith, L. V., and Yamagata, T. (2009), “A Panel Unit Root Test in the Presence of a Multifactor Error Structure,” unpublished manuscript. [114,116]

Poon, S.-H., and Granger, C. W. J. (2003), “Forecasting Volatility in Financial Markets: A Review,” *Journal of Economic Literature*, 41, 478–559. [112]

Pötscher, B. M. (1989), “Model Selection Under Nonstationarity: Autoregressive Models and Stochastic Linear Regression Models,” *The Annals of Statistics*, 17, 1257–1274. [118]

Schmidt, P., and Phillips, P. C. B. (1992), “LM Tests for a Unit Root in the Presence of Deterministic Trends,” *Oxford Bulletin of Economics and Statistics*, 54, 257–287. [114]

Sensier, M., and van Dijk, D. (2004), “Testing for Volatility Changes in US Macroeconomic Time Series,” *Review of Economics and Statistics*, 86, 833–839. [112]

Seo, B. (1999), “Distribution Theory for Unit Root Tests With Conditional Heteroskedasticity,” *Journal of Econometrics*, 91, 113–144. [112]

Shin, D. W., and Kang, S. (2006), “An Instrumental Variable Approach for Panel Unit Root Tests Under Cross-Sectional Dependence,” *Journal of Econometrics*, 134, 215–234. [119]

Warnock, M. V. C., and Warnock, F. E. (2000), “The Declining Volatility of U.S. Unemployment: Was Arthur Burns Right?,” *International Finance Discussion Article 677*, Board of Governors of the Federal Reserve System. [112]

Watson, M. W. (1999), “Explaining the Increased Variability in Long-Term Interest Rates,” *Federal Reserve Bank of Richmond, Economic Quarterly*, 85, 71–96. [112]

Westerland, J., and Breitung, J. (2013), “Lessons From a Decade of IPS and LLC,” Forthcoming in *Econometric Reviews*, 32, 547–591. [116]