Financial Econometrics Series

SWP 2014/02

HETEROSKEDASTICITY ROBUST PANEL UNIT ROOT TESTS

J. Westerlund

The working papers are a series of manuscripts in their draft form. Please do not quote without obtaining the author’s consent as these works are in their draft form. The views expressed in this paper are those of the author and not necessarily endorsed by the School or IBISWorld Pty Ltd.
Heteroskedasticity Robust Panel Unit Root Tests∗

Joakim Westerlund†
Deakin University
Australia

September 12, 2013

Abstract

This paper proposes new unit root tests for panels where the errors may be not only serial and/or cross-correlated, but also unconditionally heteroskedastic. 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.

JEL Classification: C13; C33.

Keywords: Unit root test; Panel data; Unconditional heteroskedasticity; GARCH; Cross-section dependence, Common factors.

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

∗A previous version of this paper was presented at a workshop at Deakin University. The author would like to thank seminar participants, and in particular Christoph Hanck, Rolf Larsson, Paresh Narayan, Stephan Smeekes, Jean-Pierre Urbain, Shaehee Khan (the Editor), an Associate Editor, and two anonymous referees for many valuable comments and suggestions. Financial support from the Jan Wallander and Tom Hedelius Foundation under research grant number P2009–0189:1 is gratefully acknowledged.

†Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia. Telephone: +61 3 924 46973. Fax: +61 3 924 46283. E-mail address: j.westerlund@deakin.edu.au.
of the efficient market hypothesis, requiring that stock returns are stationary, in which case violations of the otherwise so common homoskedasticity assumption is more of a rule rather than the exception. In fact, as Poon and Granger (2003, page 481) state 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, for example, 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), Loretan and Phillips (1994), Watson (1999), and Busetti and Taylor (2003), to mention a few, all provide strong evidence against the unconditional homoskedasticity assumption for most financial variables, including exchange rates, interest rates and stock returns.

Of course, unconditional heteroskedasticity 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 evidence how the volatility of US gross domestic product (GDP) has declined since the early 1980s. Busetti and Taylor (2003) consider 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 have been reported by Cavaliere and Taylor (2009a). Chauvet and Potter (2001) report 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) consider 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 paper can be seen as a response to the above observations. The purpose is to device 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) heteroskedasticity. 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 Cavaliere and Taylor, 2009b; Beare, 2008). 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 heteroskedasticity.

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 heteroskedasticity is needed, and still the implementation is extremely simple, requiring only minimal corrections. The asymptotic analysis reveals that, while the test statistics have asymptotic distributions that are free of nuisance parameters under the unit root null, this is not the case under the local alternative, in which power generally depends on both the heteroskedasticity and the cross-correlation. Results from a small Monte Carlo study show that the asymptotic results are borne out well in small samples. In fact, the tests have excellent small-sample properties, even when compared to the competition.

The rest of the paper 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.
2 Model and assumptions

Consider the panel data variable $y_{i,t}$, where $t = 1, ..., T$ and $i = 1, ..., N$ index the time series and cross-sectional units, respectively. The data generating process (DGP) of this variable is given by

\[ y_{i,t} = \alpha_i + u_{i,t}, \]
\[ u_{i,t} = \rho_i u_{i,t-1} + v_{i,t}, \]

with $v_{i,t}$ following a stationary autoregressive process of known order $p_i < \infty$ with possibly cross-section dependent and/or group-wise heteroskedastic errors. In particular, it is assumed that $v_{i,t}$ admits to the following general dynamic common factor model:

\[ \phi_i(L)v_{i,t} = e_{i,t}, \]
\[ e_{i,t} = \Theta_i'F_t + \epsilon_{i,t}, \]
\[ \epsilon_{i,t} = \sigma_{m,t}\epsilon_{i,t}, \]

where $\phi_i(L) = 1 - \sum_{k=1}^{p_i} \phi_{ik}L^k$ is a polynomial in the lag operator $L$, $F_t$ is an $r$-dimensional vector of common factors, and $m = 1, ..., M+1$ indexes $M+1$ distinct variance groups. As with $p_i$, initially we will assume that $r$ (the number of common factors) and $M$ (the number of variance groups) are known; some possibilities for how to relax these assumptions in practice will be discussed in Section 3.3. In the assumptions that follow \[ \lfloor x \rfloor \] signifies the integer part of $x$ and $\| A \| = \sqrt{\text{tr}(A' A)}$ is the Frobenius (Euclidean) norm of the matrix $A$.

Assumption 1.

(a) $\epsilon_{i,t}$ is independent and identically distributed (iid) across both $i$ and $t$ with $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = 1$ and $E(\epsilon_{i,t}^4) < \infty$;

(b) $F_t$ is iid across $t$ with $E(F_t) = 0$, $E(F_t' F_t') = \Sigma_t > 0$ and $E(||F_t||^4) < \infty$;

(c) $\epsilon_{i,t}$ and $F_s$ are mutually independent for all $i$, $t$ and $s$;

(d) $||\Theta_i|| < \infty$ for all $i$ and $\sum_{i=1}^{N} \Theta_i' \Theta_i' / N \to \Lambda > 0$ as $N \to \infty$;

(e) $\phi_i(L)$ have all its roots outside the unit circle;

(f) $E(u_{i,s}) < \infty$ for all $i$ and $s = -p_i, ..., 0$. 

4
Consider $\sigma^2_{m,t}$. It is assumed that the cross-section can be divided into $M$ cross-section homoskedastic groups with the $m$-th group containing the units $i = N_{m-1} + 1, ..., N_m$, where $N_0 = 1$ and $N_{M+1} = N$. That is, the heteroskedasticity across the cross-section is made up of $M$ discrete jumps between otherwise homoskedastic 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 $i = N_{m-1} + 1, ..., N_m$ 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 $q > 0$, $\lambda_0 < \lambda_1 < ... < \lambda_M < \lambda_{M+1}$, $\lambda_0 = 0$ and $\lambda_{M+1} = 1$;

(b) $\sigma^2_{m,[sT]} \to \sigma_m^2(s) > 0$ and $\Sigma_{[sT]} \to \Sigma(s) > 0$ as $T \to \infty$, where $\sigma_m^2(s)$ and $\Sigma(s)$ are bounded and square-integrable with a finite number of points of discontinuity;

(c) $\sum_{i=1}^{N} \sigma^2_{i,[sT]} \Theta_i \Theta_i' / N \to \Lambda(s) > 0$ as $N, T \to \infty$.

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 $\epsilon_{i,t}$ and $F_s$ for all $i, j, t$ and $s$.

Hence, if $c_i = 0$, then $\rho_i = 1$, and therefore $y_{i,t}$ is unit root non-stationary, whereas if $c_i < 0 (c_i > 0)$, then $\rho_i$ approaches one from below (above) and therefore $y_{i,t}$ is “locally” stationary (explosive) (in the mean). The relevant null hypothesis here is given by $H_0 : c_1 = ... = c_N = 0$, 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 parametrization, which we will be working under when studying the local asymptotic power of our time series statistic (see Section 3.1). 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 non-negligible 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 \( \epsilon_{i,t} \) is key in our approach. The assumption of “group-wise heteroskedasticity” in the cross-section 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 is 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 paper).

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

3. The assumption in (4) is quite common (see, for example, Moon and Perron, 2004; Pesaran, 2007; Pesaran et al., 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 (2004, 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 (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 $\text{var}(c_i)$ have limits such as $\mu_c$ and $\sigma_c^2$, respectively. The assumption that $c_i$ is independent of $\varepsilon_{jt}$ and $F_s$ can also be relaxed somewhat. In particular, note that since under the null of a unit root $c_1 = \ldots = c_N = 0$, correlation between $c_i$ and $\varepsilon_{jt}$ and/or $F_s$ is only an issue under the alternative hypothesis, and then it can be shown that in the typical panel parametrization of $\rho_i$ with $\kappa = 1/2$, the effect of correlation between $c_i$ and $\varepsilon_{jt}$ 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 $\tau_i$, is very simple and can be seen as a version of the Lagrange multiplier (LM) test statistic for a unit root (see Remark 4 below). 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 $(\Delta y_{i,t-1}, \ldots, \Delta y_{i,t-p_i})$.

2. Obtain $\hat{F}_t$ and $\hat{\Theta}_i$ by applying the principal components method to $\Delta \hat{R}_{i,t}$. Compute $\hat{\epsilon}_{i,t} = \Delta \hat{R}_{i,t} - \hat{\Theta}_i \hat{F}_t$.

3. Compute $\hat{w}_{m,t} = 1/\hat{\sigma}_{m,t}$, where $\hat{\sigma}_{m,t}^2 = \sum_{i}^{N_m} \frac{\epsilon_{i,t}^2}{N_m - N_{m-1}}$.

---

1Because $p_i$ is not restricted to be equal across the cross-section, the panel data variable $\hat{\epsilon}_{i,t}$ will in general be unbalanced. Therefore, in order to simplify this step of the implementation, one may apply the principal components method to the balanced panel comprising the last $T - \max\{p_1, \ldots, p_N\} - 1$ observations.
4. Compute

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

where $i = N_{m-1} + 1, ..., N_m$, $\Delta \hat{R}_{wi,t} = \hat{w}_{m,t} \hat{e}_{i,t}$ and $\hat{R}_{wi,t} = \sum_{k=p_i+2}^{t} \Delta \hat{R}_{wi,k}$.

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

$$
\frac{\left( \sum_{t=p_i+2}^{T} w_{m,t}^2 \Delta R_{i,t} R_{i,t-1} \right)^2}{\sum_{t=p_i+2}^{T} w_{m,t}^2 R_{i,t-1}^2},
$$

where $w_{m,t} = 1/\sigma_m(L)$, $\Delta R_{i,t} = \phi(L) \Delta y_{i,t}$ and $R_{i,t} = \sum_{k=p_i+2}^{t} \Delta R_{i,k}$. 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 $w_{m,t} R_{i,t-1}$ with $\sum_{s=p_i+2}^{t} w_{m,s} \Delta R_{i,s}$) 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 conventional ADF test. It can also be regarded as a generalization of the LM test developed by Ahn (1993), and Schmidt and Phillips (1992), who consider the problem of testing for a unit root in a pure time series setting with homoskedastic errors. The way that $\Delta \hat{R}_{wi,t}$ is accumulated up to levels is very similar to the approach of Bai and Ng (2004). But while in that paper 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 heteroskedasticity (see Beare, 2008, for a similar approach).

7. As is well-known, under certain conditions (such as normality), the LM test statistic possess 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 $\tau_i$.

\footnote{A formal derivation is available upon request.}

\footnote{Another problem with the true LM test statistic is that the feasible version is based on using weighted OLS (WLS) to estimate the first-step regression. However, since the weight $w_{m,t}$ is not known, this calls for the use of iterated WLS.}
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 \( y_{t,i} \) unit root non-stationary and proceed no further, whereas if \( H_0 \) is rejected, then the testing continues by checking the sign of \( \sum_{t=p+2}^{T} \Delta \hat{R}_{wi,t} \hat{R}_{wi,t-1} \). 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 \( 1(\sum_{t=p+2}^{T} \Delta \hat{R}_{wi,t} \hat{R}_{wi,t-1} / T < 0) \tau_i \), 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 heteroskedastic errors it is common practice to use \( \hat{\sigma}^2_{m,t} \) to estimate \( \sigma^2_{m,t} \). However, as Lopez (2001) points out, while unbiased, because of its asymmetric distribution, \( \hat{\sigma}^2_{m,t} \) is a very imprecise estimator of \( \sigma^2_{m,t} \). \( \hat{\sigma}^2_{t,m} \) uses more information and is therefore expected to perform much better. Beare (2008) uses a similar recursive weighting scheme as the one considered here, but where \( \sigma^2_{m,t} \) 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 \( N, T \to \infty \)

\[
\tau_i \to_d \frac{\left( \int_{s=0}^{1} V_{R,J}(s) dV_{R,J}(s) \right)^2}{\int_{s=0}^{1} V_{R,J}(s)^2 ds},
\]
where

\[
\begin{align*}
    w_m(s) &= \frac{1}{\sigma_m(s)}, \\
    V_{R,i}(s) &= W_{\epsilon,i}(s) + c_i \int_{t=0}^{s} w_m(r) B_{R,i}(r)dr, \\
    B_{R,i}(s) &= \Theta_i' B_{F,i}(s) + B_{\epsilon,i}(s), \\
    B_{\epsilon,i}(s) &= \int_{t=0}^{s} \exp((s-r)c_i)\sigma_m(r)dw_{\epsilon,i}(r), \\
    B_{F,i}(s) &= \int_{t=0}^{s} \exp((s-r)c_i)\Sigma(r)^{1/2}dw_F(r),
\end{align*}
\]

with \( \rightarrow_d \) signifying convergence in distribution, \( i = N_m - 1, \ldots, N_m \), and \( W_{\epsilon,i}(s) \) and \( W_F(s) \) being two independent standard Brownian motions.

In order to appreciate fully the implication of Theorem 1, note that since \( V_{R,i}(s) \) satisfies the differential equation \( dV_{R,i}(s) = dw_{\epsilon,i}(s) + c_i w_m(s) B_{R,i}(s)ds \), the numerator of the asymptotic test distribution can be expanded as

\[
\int_{s=0}^{1} V_{R,i}(s)dV_{R,i}(s) = c_i \int_{s=0}^{1} V_{R,i}(s)w_m(s)B_{R,i}(s)ds + \int_{s=0}^{1} V_{R,i}(s)dw_{\epsilon,i}(s).
\]

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 heteroskedasticity 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 homoskedastic 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 heteroskedasticity.

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

\[
\tau_i \rightarrow_d \frac{\left( \int_{s=0}^{1} W_{\epsilon,i}(s)dw_{\epsilon,i}(s) \right)^2}{\int_{s=0}^{1} W_{\epsilon,i}(s)^2ds}.
\]
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, for example, Ahn (1993), who also tabulate critical values (Table 1).

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 heteroskedasticity of the type considered here (but no cross-section dependence). However, unlike $\tau_i$, the asymptotic distribution of this test statistic depends on $\sigma_m^2(r)$, which is of course never known in practice. Similarly, while the unit root tests of Pesaran (2007) and Pesaran et al. (2009) do allow for cross-section dependence in the form of common factors (but no heteroskedasticity), 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 heteroskedasticity 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 estimation introduces a dependence on $\sigma_m^2(s)$. One way to circumvent this problem is to use bootstrap techniques (see, for example, Cavaliere and Taylor, 2009b). However, this does not fit well with the simple and parametric flavor of our approach. A feasible alternative in cases with trending data is to assume a common trend slope, which can then be removed by using data that have been demeaned with respect to the overall sample mean. That is, instead of using $\Delta y_{i,t}$ when constructing $\Delta \hat{R}_{wi,t}$ one uses $\Delta y_{i,t} - \bar{y}$, where $\bar{y} = \frac{1}{N} \sum_{j=1}^{N} \sum_{k=p_i+2}^{T} \Delta y_{jk}/NT$. The asymptotic distribution reported in Theorem 1 is unaffected by this.\footnote{A detailed proof is available upon request.} Some limited heterogeneity can also be permitted.
in this way by assuming that the trend slope, \( \beta_i \) say, is “local-to-constant”. Specifically, supposed that instead of (1), we have

\[
y_{i,t} = \alpha_i + \beta_i t + u_{i,t},
\]

where \( \beta_i = \beta + b_i / T^\gamma, \gamma > 1 \) and \( b_i \) is iid across \( i \) with \( E(b_i) = 0 \) and \( \text{var}(b_i) = \sigma^2_b < \infty \). \( b_i \) is independent of all the other random elements of the DGP. The results reported in Theorem 1 (and also Theorem 2) hold also in this case, provided again that \( \Delta y_{i,t} \) is replaced with \( \Delta y_{i,t} - \overline{\Delta y} \).\(^5\)

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 homoskedastic 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 \( \tau_i \):

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

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 \( N, T \to \infty \) with \( N/T \to 0 \)

\[
\tau \to_d \chi^2_1(\theta^2),
\]

\(^5\)Some confirmatory Monte Carlo results are available upon request.
where

\[
\theta = \lim_{N \to \infty} \frac{\sum_{m=1}^{M+1} (\lambda_m - \mu_m) \int_0^1 q_{N,m}(r) dr}{\sqrt{\sum_{m=1}^{M+1} (\lambda_m - \mu_m) \int_0^1 \overline{p}_{N,m}(r) dr}},
\]

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

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

\[
p_{N,i}(r) = r + \frac{2 \mu_i}{\sqrt{N}} \int_{u=0}^r g_m(u) du + \left( \frac{\mu_i^2 + \sigma_i^2}{N} \right) \int_{u=0}^r \left( g_i(u, u) + 2 \int_{v=0}^u g_i(u, v) dv \right) du,
\]

\[
q_{N,i}(u) = \mu_i g_m(u) + \left( \frac{\mu_i^2 + \sigma_i^2}{N} \right) \int_{v=0}^u g_i(u, v) dv,
\]

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

\[
g_i(u, v) = w_m(u) w_m(v) \int_{x=0}^u (\Theta_i^\prime \Sigma(x) \Theta_i + \sigma_m(x)^2) dx,
\]

and $\chi^2_k(d)$ is a non-central chi-square distribution with $k$ degrees of freedom and non-centrality parameter $d$.

In agreement with the results reported in Theorem 1 the asymptotic local power of $\tau$ depends on both the heteroskedastisity 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.\(^6\)

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 $N \to \infty$). 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.

In order to illustrate how a particular choice of $\sigma^2_m(s)$ 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,

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

\(^6\)Demetrescu and Hanck (2011) considers a similar setup but in order to work out the local power they assume that $\Theta_i = 0$, thereby disregarding the power effect of the cross-section dependence.
suggesting that $\int_{r=0}^{1} \overline{p}_{N,m}(r) dr \to \mu_c(1/2 + 3b(1-b))$ as $N \to \infty$. But we also have that
$\int_{r=0}^{1} \overline{p}_{N,m}(r) dr \to \int_{r=0}^{1} r dr = 1/2$, 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 heteroskedasticity ($b \in (0,1)$) than without ($b = 0$ or $b = 1$). The maximal value of $\theta^2$ is obtained by setting $b = 1/2$.

Under the null hypothesis that $c_i$ and all its moments are zero, $\overline{q}_{N,m}(u) = 0$, suggesting that the result in Theorem 2 reduces to
$$\tau \to_d \chi^2_1(0) \sim \chi^2_1,$$
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 $N \to \infty$ (and also $T \to \infty$) and only report results for the resulting first-order approximate power function (see, for example, Moon et al., 2007), which only depends on $\mu_c$, Theorem 2 also accounts for $\sigma^2_c$, 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 $\sigma^2_c$ goes to zero as $N \to \infty$.

13. While non-negligible 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 for $\tau_i$ to be able to detect them. The fact that $\tau$ has non-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 $N/T \to 0$ as $N, T \to \infty$ is standard even when testing for unit roots in homoskedastic 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 $N/T \to 0$ as $N, T \to \infty$ prevents this error from having a dominating effect (see Westerlund and Breitung, 2011, for a detailed discussion). Demetrescu and Hanck (2011) considers a setup that is similar to ours. However, they require $N/T^{1/5} \to 0$, which obviously puts a limit on the allowable sample sizes. For example, if $N = 10$, then $T > 100,000$ is required for $T > N^5$ 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

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 group-wise 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 group-wise 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 $M + 1 = 2$ groups. The problem of determining the groups therefore reduces to the problem of consistent estimation of the group threshold $N_1$. For this purpose it is convenient to treat $\hat{\sigma}^2_{1,t}$ and $\hat{\sigma}^2_{2,t}$ as functions of $n = qN, \ldots, (1-q)N$, that is, $\hat{\sigma}^2_{1,t}(n) = \sum_{i=1}^{n} \hat{\epsilon}^2_{i,t}/n$ and $\hat{\sigma}^2_{2,t}(n) = \sum_{i=n+1}^{N} \hat{\epsilon}^2_{i,t}/(N-n)$. In this notation, the QML objective function is given by

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

and the proposed estimator of $N_1$ 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 $N_1$. 

15
Proposition 1. Under the conditions of Theorem 1, given $\kappa \geq 0$, as $N, T \to \infty$ with $N/T \to 0$

$$P(\hat{N}_1 = N_1) \to 1.$$ 

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 $\hat{\sigma}_i^2 = \sum_{t=p_i+2}^{T} \hat{\epsilon}_t^2 / T$ 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 $N_1$; 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).\footnote{Bai (1997) suggests 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.}

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 $\hat{\sigma}_i^2$ 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 heteroskedasticity in $t$ (see Lin and Ng, 2012).

16. An alternative to the above break estimation approach to the grouping is to use $K$-means clustering (see, for example, 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.

Lag selection

It is well known that if the true lag order $p_i$ is unknown but the estimator $\hat{p}_i$ is such that $\hat{p}_i \to \infty$ and $\hat{p}_i = o(T^{1/3})$, 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 $p_i$ is infinite and the errors are (unconditionally) heteroskedastic (Cavaliere and Taylor, 2009b). One possibility is therefore to set $\hat{p}_i$ as a function of $T$. Of course, in practice it is often preferable to use a data-driven approach. In this context, Pötscher (1989) shows that if the errors are unconditionally heteroskedastic, 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 heteroskedasticity 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:

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

where $\tau_i(p) = \hat{\gamma}_i^2(p) \sum_{t=p_{\max}+2}^{T} \hat{R}_{i,t-1}^2 / \hat{\sigma}^2_i(p)$, $\hat{\gamma}_i(p)$ 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 $p_{\max}$ and is assumed to
satisfy $p_{\text{max}} \to \infty$ and $p_{\text{max}} = o(T^{1/3})$. The estimator of $p_i$ is given by

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

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

### Selection of the number of common factors

In 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 heteroskedasticity. Therefore, if the number of common factors, $r$, is unknown, as is typically the case in practice, any of the information criteria proposed by Bai and Ng (2002) can be applied to $\Delta \hat{R}_i$, giving $\hat{r}$

If $p_i$, $r$ 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 $p_{\text{max}}$. Once $r$ has been estimated, the MBIC is applied to obtain $\hat{p}_i$ for each unit. Given $\hat{r}$ and $\hat{p}_1, \ldots, \hat{p}_N$, the groups are selected by applying QML to $\hat{\epsilon}_{i,t}$.

### 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 $\alpha_i = 1$, $u_{i,0} = 0$, $\phi_i(L) = 1 - \phi_1 L$, $\Theta_i \sim N(1,1)$, $F_t \sim N(0,1)$, $\epsilon_{i,t} \sim N(0,1)$ and $c_i \sim U(a,b)$. Three cases regarding the time-variation in $\sigma_{m,t}^2$ are considered: (1) $\sigma_t^2 = 1$; (2) $\sigma_t^2 = 1 - 1/(t \geq \lfloor T/2 \rfloor 3/4)$; (3) $\sigma_t^2 = 1 - 3/4(1 + \exp(-(t - \lfloor T/2 \rfloor)))$. While case 2 corresponds to a single discrete break at time $\lfloor T/2 \rfloor$ when $\sigma_t^2$ changes from 1 to 1/4, case 3 corresponds to a smooth transition break from 1 to 1/4 with $\lfloor T/2 \rfloor$ 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 $\sigma_{m,t}^2 = \sigma_t^2(1 + 1(i > \lfloor N/2 \rfloor))$. Hence, in this DGP there are two groups. In the first, containing the first $\lfloor N/2 \rfloor$ units, $\sigma_{m,t}^2 = \sigma_t^2$, while in the second, containing the remaining units, $\sigma_{m,t}^2 = 2\sigma_t^2$. 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 $\tau_{IV,j}$ and $\tau_{IV}$ test statistics of Demetrescu and Hanck
Both 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 $\tau_{IV,i}$, one 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 heteroskedasticity (see Demetrescu and Hanck, 2011, Proposition 1), the sign-based $\hat{S}_{NT}$ statistic of Shin et al. (2009), which in construction is very close to $\tau_{IV}$, is also simulated. A number of other tests were attempted (including the $ADF_c(i)$ and $P_c^c$ statistics of Bai and Ng, 2004) but not included, as their performance was dominated by the performance of the above tests.\(^8\)

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, for example, 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 $\sigma^2_t$.

- The most serious distortions are found for $\tau_{IV}$, which is generally quite severely under-

---

\(^8\)Following the recommendation of Demetrescu and Hanck (2011), two versions of $\tau_{IV}$ were considered. The first is the original test statistic also considered by Shin and Kang (2006), in which the data are weighted by the inverse of the conventional estimator of the cross-sectional covariance matrix of the errors. The second test is based on a modified covariance matrix estimator that supposed to work better in small samples. The results were, however, very similar, and we therefore only report the results from the latter test. Also, the test that we denote here by $\tau_{IV,i}$ is the weighted time series test that is used in constructing $\tau_{IV}$. While Demetrescu and Hanck (2011) do not present any asymptotic results for this test, in view of their Proposition 2, it seems reasonable to expect it to share the robustness of $\tau_{IV}$.

\(^9\)See Demetrescu and Hanck (2011, Section 4) for a comprehensive Monte Carlo comparison with their tests.
sized, especially when $\phi_1 = 0.5$ 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 $N, T \to \infty$) is not even close to being satisfied, and therefore the poor performance of $\bar{\tau}_{IV}$ 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. $\tau_{IV,j}$ is also undersized, although not to the same extent as $\bar{\tau}_{IV}$.

- 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 should in principle apply to $\bar{\tau}_{IV}$. 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 $T < N^5$. The choice $\kappa = 1/2$ is similarly too fast for $\tau_i$ and $\tau_{IV,i}$ to have non-negligible 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, $\tau_{IV,j}$ and $\bar{\tau}_{IV}$ 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 $\bar{\tau}_{IV}$!

- Power depends on the nature of the heteroskedasticity. 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 $\Delta \hat{R}_{wi,t} = \hat{w}_{i,t} \hat{\epsilon}_{i,t} = \text{sign} (\hat{\epsilon}_{i,t})$, and sign-based tests have been shown to be robust to certain types of heteroskedasticity (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 paper 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 heteroskedasticity. In the paper we assume that the heteroskedasticity is deterministic, as when considering models of permanent shifts or linear time trends, but it could also be stochastic. Conditional heteroskedasticity in the form of, for example, GARCH is permitted too. Moreover, the heteroskedasticity is not restricted to time series dimension, but can also emanate to some extent from the cross-sectional dimension. The assumed DGP is therefore very general, and much more so than in most other panel data studies. In fact, the only other panel study that we are aware of to consider such a general setup is that of Demetrescu and Hanck (2011).

Two tests are proposed. One is designed to test for a unit root in a single cross-section unit, while the other is designed to test for a unit root in the whole panel. Both tests are remarkably simple, requiring no prior knowledge regarding the heteroskedasticity, and still only minor corrections are needed in order to obtain asymptotically nuisance parameter free test distributions. This invariance is verified in small samples using Monte Carlo simulation. It is found that the new tests show smaller size 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.
References


Westerlund, J., and J. Breitung (2012). Lessons from a Decade of IPS and LLC. Forthcoming in *Econometric Reviews*.
Appendix: Proofs

Lemma A.1. Under Assumptions 1–3,

\[
\Delta \hat{R}_{wi,t} = \Delta R_{wi,t} + \mathcal{G}_{i,t} + O_p \left( \frac{1}{\sqrt{NT}} \right),
\]

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

where

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

\[
R_{wi,t} = \sum_{s=p_i+2}^{l} \epsilon_{i,t} + (\rho_i - 1) \sum_{s=p_i+2}^{l} w_{m,s}R_{i,s-1},
\]

\[
\mathcal{G}_{i,t} = \Theta_i' H^{-1} w_{m,t}v_t - \frac{1}{2} w_{m,t}^T (\mathcal{C}_{0m,t} + \mathcal{C}_{4m,t} + \mathcal{C}_{6m,t}) \epsilon_i, t
\]

\[
G_{i,t} = \Theta_i' H^{-1} \sum_{s=p_i+2}^{l} w_{m,s} v_s - \frac{1}{2} \sum_{s=p_i+2}^{l} w_{m,s} \mathcal{C}_{0m,s} \epsilon_i, s,
\]

\[
\mathcal{T}_{m,t} = \frac{1}{(N_m - N_{m-1})} \sum_{i=N_{m-1}+1}^{N_m} \mathcal{C}_{6i,t},
\]

with \( i = N_{m-1} + 1, \ldots, N_m \), \( R_{i,t} = \phi_i(L)u_i,t \), \( c_{0i,t} = (\epsilon_{i,t}^2 - \sigma_{m,t}^2) \), \( c_{4i,t} = 2\epsilon_{i,t} a_{i,t} \), \( c_{6i,t} = -2(\Phi_i - \Phi_i')' \epsilon_{i,t} d_{i,t} \), \( a_{i,t} = \Theta_i' H^{-1} v_t - d_i' \hat{F}_{i,t} \), \( v_t = (\hat{F}_{i,t} - HF_{i}) \), \( d_i = (\hat{\Theta}_i - (H^{-1})' \Theta_i) \), \( H \) is a \( r \times r \) positive definite matrix and \( C_{2NT} = \min\{N, \sqrt{T}\} \).

Proof of Lemma A.1.

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

\[
R_{i,t} = \phi_i(L)u_{i,t} = \phi_i(L)(y_{i,t} \cdot \alpha_i) = y_{i,t} - \Phi_i x_{i,t-1} - \phi_i(1) \alpha_i,
\]

whose first difference is given by

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

It follows that

\[
\Delta R_{i,t} = (\rho_i - 1) R_{i,t-1} + \epsilon_{i,t}, \quad (A1)
\]

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

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

\[
\Delta \hat{R}_{i,t} = \Delta y_{i,t} - \Phi_i' \Delta x_{i,t-1} = \Delta R_{i,t} - (\Phi_i - \Phi_i)' \Delta x_{i,t-1}.
\]
Note that by Taylor expansion of \( \exp(x) \) about \( x = 0, \exp(x) = 1 + x + o(1) \), suggesting that 
\[
\rho_i = \exp(c_i / N^\kappa T) = 1 + c_i / N^\kappa T + o_p(1).
\]
This implies 
\[
\frac{1}{\sqrt{T}} \sum_{t=p_i+2}^T \Delta x_{i,t-1} \Delta R_{i,t} = \frac{c_i}{N^\kappa T^{3/2}} \sum_{t=p_i+2}^T \Delta x_{i,t-1} R_{i,t-1} + \frac{1}{\sqrt{T}} \sum_{t=p_i+2}^T \Delta x_{i,t-1} e_{i,t} + o_p(1).
\]
In the proof of Theorem 1, we show that \( R_{i,t-1} / \sqrt{T} \) satisfies an invariance principle, and therefore that the first term on the right hand side is \( O_p(1/N^\kappa \sqrt{T}) \). As for the second term, by the Wold decomposition, \( 1/\phi(L) \) can be expanded as \( \phi^*(L) = 1/\phi(L) = \sum_{j=0}^\infty \phi_j^* L^j \) with \( \phi_0^* = 1 \). It follows that, \( \Delta y_{i,t} = (\rho_i - 1)(y_{i,t} - \alpha_i) + v_{i,t}, \) where \( v_{i,t} = \phi^*(L)e_{i,t} \). Hence, since \( e_{i,t} \) is serially uncorrelated, we have that \( E(e_{i,s} \Delta x_{k,s-1}) = 0 \) for all \( i, k \) and \( s \). The variance of \( \sum_{t=p_i+2}^T \Delta x_{i,t-1} e_{i,t} / \sqrt{T} \) is bounded. Hence, \( \sum_{t=p_i+2}^T \Delta x_{i,t-1} e_{i,t} / \sqrt{T} = O_p(1) \). Subsequently, using \( \hat{\Phi}_i \) to denote the OLS estimator of \( \Phi_i \),
\[
\sqrt{T}(\hat{\Phi}_i - \Phi_i) = \left( \frac{1}{T} \sum_{t=p_i+2}^T \Delta x_{i,t-1} (\Delta x_{i,t-1})' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=p_i+2}^T \Delta x_{i,t-1} \Delta R_{i,t} \\
= \left( \frac{1}{T} \sum_{t=p_i+2}^T \Delta x_{i,t-1} (\Delta x_{i,t-1})' \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=p_i+2}^T \Delta x_{i,t-1} e_{i,t} + o_p(1) = O_p(1).
\]
The presence of the factors therefore does not affect the rate of consistency of \( \hat{\Phi}_i \). Making use of this result, we obtain
\[
\Delta \hat{R}_{i,t} = \Delta y_{i,t} - \hat{\Phi}_i' \Delta x_{i,t-1} = \Delta R_{i,t} - (\hat{\Phi}_i - \Phi_i)' \Delta x_{i,t-1} = \Delta R_{i,t} + o_p(\frac{1}{\sqrt{T}}), \quad (A2)
\]
and therefore, with \( R_{i,t-1} = O_p(\sqrt{T}) \) and \( (\rho_i - 1) = O_p(1/N^\kappa T) \),
\[
\Delta \hat{R}_{i,t} = \Delta R_{i,t} + o_p(\frac{1}{\sqrt{T}}) = e_{i,t} + (\rho_i - 1) R_{i,t-1} + o_p(\frac{1}{\sqrt{T}}) = e_{i,t} + o_p(\frac{1}{\sqrt{T}}). \quad (A3)
\]
Since \( e_{i,t} = \Theta_i F_t + e_{i,t} \), we have that the error incurred by applying the principal components method to \( \Delta \hat{R}_{i,t} \) rather than to \( e_{i,t} \) is negligible (a detailed proof is available upon request). In order to also show that the estimation is unaffected by the heteroskedasticity in \( e_{i,t} \), we verify Assumption C in Bai (2003). Towards this end, note that, since \( N_m = \lfloor \lambda_m N \rfloor \),
\[
\frac{1}{N_m} \sum_{i=1}^{N_m} E(e_{i,t}^2) = \sum_{m=1}^{M+1} \frac{(N_m - N_{m-1})}{N} \sum_{i=N_{m-1}+1}^{N_m} E(e_{i,t}^2) \\
= \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \frac{1}{(N_m - N_{m-1})} \sum_{i=N_{m-1}+1}^{N_m} E(e_{i,t}^2) \\
= \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \sigma_{m,t}^2.
\]
which is bounded under our assumptions, and therefore so is $\sum_{i=1}^{N} \sum_{j=1}^{N} E(\epsilon_{i,t} \epsilon_{j,t}) = \sum_{i=1}^{N} E(\epsilon_{i,t}^2) / N$. Similarly,

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

which is bounded too, because $\sum_{t=1}^{T} \sigma_{m,t}^2 / T \rightarrow \int_{r=0}^{1} \sigma_{m}(r) dr < \infty$ as $T \rightarrow \infty$. Similarly, since $\sum_{t=1}^{T} F_t \sigma_{m,t}^2 / T$ 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 $\Theta'_t$ and $F_t$, respectively, are unaffected by the hetoskedasticity.

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

$$\hat{\epsilon}_{i,t} = \hat{\epsilon}_{i} - \hat{\Theta}'_{t} \hat{F}_{t} = \epsilon_{i,t} + (\rho_i - 1) R_{i,t-1} + a_{i,t} - (\Phi_i - \Phi_t)' \Delta x_{i,t-1},$$

(A4)

where, following Bai and Ng (2004, page 1154),

$$a_{i,t} = \Theta'_t F_t - \Theta'_t \hat{F}_t = \Theta'_t H^{-1}(\hat{F}_t - HF_t) - (\hat{\Theta}_i - (H^{-1})' \hat{\Theta}_i)' \hat{F}_t = \Theta'_t H^{-1} \nu_t - d'_i \hat{F}_t,$$

with $\nu_t$ and $d_t$ implicitly defined. Hence,

$$\hat{\epsilon}_{i,t}^2 = (\epsilon_{i,t} + (\rho_i - 1) R_{i,t-1} + a_{i,t} - (\Phi_i - \Phi_t)' \Delta x_{i,t-1})^2 = \epsilon_{i,t}^2 + c_{1i,t} + ... + c_{9i,t},$$

(A5)

where

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

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

$$c_{3i,t} = (\Phi_i - \Phi_t)' \Delta x_{i,t-1} \Delta x_{i,t-1}' (\Phi_i - \Phi_t),$$

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

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

$$c_{6i,t} = -2 (\Phi_i - \Phi_t)' \epsilon_{i,t} \Delta x_{i,t-1},$$

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

$$c_{8i,t} = -2 (\Phi_i - \Phi_t)' a_{i,t} \Delta x_{i,t-1},$$

$$c_{9i,t} = -2 (\rho_i - 1) (\Phi_i - \Phi_t)' R_{i,t-1} \Delta x_{i,t-1}.$$

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 $\hat{\delta}_{m,t}^2$, which is $\hat{\sigma}_{m,t}^2$. 

29
for the first group. Letting $c_{0,l,t} = (e_{1,t}^2 - a_{1,t}^2)$ and using $c_{si,l,t} = \sum_{i=1}^{N_i} c_{si,l,t} / N_1$ to denote the average $c_{si,l,t}$ for this group, we have

$$\delta_{1,l,t}^2 = \frac{1}{N_1} \sum_{i=1}^{N_i} \delta_{i,t}^2 = \frac{1}{N_1} \sum_{i=1}^{N_i} e_{1,t}^2 + \tau_{11,l + \ldots + \tau_{99,l} = \delta_{1,t}^2 + \tau_{01,l} + \ldots + \tau_{99,l} = \delta_{1,t}^2 + \bar{C}_{1,l,t},} \tag{A6}$$

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

$$\bar{w}_{1,l} = \frac{1}{\bar{\delta}_{1,l}}$$

$$= w_{1,l} - \frac{1}{2} w_{1,t} (\delta_{1,t}^2 - \sigma_{1,t}^2) + \frac{3}{4} w_{1,t} (\delta_{1,t}^2 - \sigma_{1,t}^2)^2 + O_p((\delta_{1,t}^2 - \sigma_{1,t}^2)^3)$$

$$= w_{1,t} - \frac{1}{2} w_{1,t} \bar{C}_{1,t} + \frac{3}{4} w_{1,t} \bar{C}_{1,t}^2 + O_p(\bar{C}_{1,t}^3). \tag{A7}$$

In order 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 $||v_t|| = O_p(1/C_{1NT})$ 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 (A7) has to be of second order, because otherwise the resulting remainder in the numerator of $\tau, \sum_{t=1}^{N_i} \sum_{t=p_i+2}^{T} \Delta \hat{R}_{wi,t} \hat{R}_{wi,t-1} / \sqrt{N_T}$, would not necessarily be negligible. For the numerator of $\tau_t, \sum_{t=p_i+2}^{T} \Delta \hat{R}_{wi,t} \hat{R}_{wi,t-1} / T$, it is enough with a first-order expansion.

We begin with $\tau_{01,l,t}$, whose order is given by

$$\tau_{01,l,t} = \frac{1}{N_1} \sum_{i=1}^{N_i} \left( e_{1,t}^2 - a_{1,t}^2 \right) = O_p \left( \frac{1}{\sqrt{N}} \right), \tag{A8}$$

as follows from using $E(e_{1,t}^2 - a_{1,t}^2) = 0$, cross-section independence and $N_1 = \lambda_1 N = O(N)$. As for $\tau_{21,t}$, by using $(a + b)^2 \leq 2(a^2 + b^2)$ and $||AB|| \leq ||A|| ||B||$ for arbitrary scalars $a$ and $b$, and conformable matrices $A$ and $B$, it is clear that $c_2_{l,t} = a_{l,t}^2 \leq 2||\Theta_i H^{-1}||^2 ||v_t||^2 + 2||d_i||^2 ||\hat{F}_i||^2$, and therefore

$$\tau_{21,l,t} = \frac{1}{N_1} \sum_{i=1}^{N_i} a_{l,t}^2 \leq \frac{2}{N_1} \sum_{i=1}^{N_i} ||\Theta_i H^{-1}||^2 ||v_t||^2 + \frac{2}{N_1} \sum_{i=1}^{N_i} ||d_i||^2 ||\hat{F}_i||^2 = O_p \left( \frac{1}{C_{2NT}} \right), \tag{A9}$$

where $C_{NT} = C_{1NT} + C_{2NT} = \min\{\sqrt{N}, \sqrt{T}\}$. Next, consider $\tau_{41,l,t}$. Because $\hat{F}_t = HF_t + v_t$, we have $\epsilon_{i,t} d_i \hat{F}_t = \epsilon_{i,t} d_i HF_t + \epsilon_{i,t} d_i v_t$, where, by the Cauchy–Schwarz inequality,

$$|| \frac{1}{N_1} \sum_{i=1}^{N_i} \epsilon_{i,t} d_i || \leq \left( \frac{1}{N_1} \sum_{i=1}^{N_i} \epsilon_{i,t}^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_i} ||d_i||^2 \right)^{1/2} = O_p \left( \frac{1}{C_{2NT}} \right).$$
It follows that
\[
\left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} d_i' HF_i \right| \leq \left( \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} d_i \right) \left| HF_i \right| = O_p \left( \frac{1}{C_{2NT}} \right),
\]
\[
\left| \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} d_i' v_i \right| \leq \left( \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} d_i \right) \left| v_i \right| = O_p \left( \frac{1}{C_{1NTC_{2NT}}}, \right),
\]
and so, via the cross-section independence of \( \epsilon_{i,j} \),
\[
|\bar{z}_{41,1}| = \left\| \frac{2}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} a_{i,t} \right\| \leq \left( \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} O_i \right) \left| H^{-1} \right| \left| v_i \right| + \left( \frac{1}{N_1} \sum_{i=1}^{N_1} \epsilon_{i,j} d_i \right) \left| \hat{f}_i \right| = O_p \left( \frac{1}{\sqrt{N_1 N C_{1NT}}} \right) + O_p \left( \frac{1}{C_{2NT}} \right). \tag{A10}
\]

Similarly, since \((\rho_i - 1) = O_p(1/N^x T)\),
\[
|\bar{z}_{71,1}| = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} (\rho_i - 1) a_{i,t} R_{i,t-1} \right| \leq 2 \left( \frac{1}{N_1} \sum_{i=1}^{N_1} a_{i,t}^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_1} (\rho_i - 1)^2 R_{i,t-1}^2 \right)^{1/2}
= O_p \left( \frac{1}{\sqrt{T N^x C_{NT}}} \right), \tag{A11}
\]
\[
|\bar{z}_{81,1}| = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} (\Phi_i - \Phi_i)^{\prime} a_{i,t} \Delta x_{i,t-1} \right|
\leq 2 \left( \frac{1}{N_1} \sum_{i=1}^{N_1} a_{i,t}^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||\Phi_i - \Phi_i||^2 ||\Delta x_{i,t-1}||^2 \right)^{1/2}
= O_p \left( \frac{1}{\sqrt{T C_{NT}}} \right). \tag{A12}
\]

The same argument can be used to show that \( \bar{z}_{11,1}, \bar{z}_{31,1} \) and \( \bar{z}_{91,1} \) are \( O_p(1/T) \). It remains to consider \( \bar{z}_{51,1} \) and \( \bar{z}_{61,1} \). The orders of these are as follow:
\[
\bar{z}_{51,1} = \frac{2}{N_1} \sum_{i=1}^{N_1} (\rho_i - 1) \epsilon_{i,j} R_{i,t-1} = \frac{2}{N_1 N^{1/2+x} T} \sum_{i=1}^{N_1} \epsilon_{i,j} R_{i,t-1} = O_p \left( \frac{1}{N^{1/2+x} \sqrt{T}} \right), \tag{A13}
\]
\[
|\bar{z}_{61,1}| = \left| \frac{2}{N_1} \sum_{i=1}^{N_1} (\Phi_i - \Phi_i)^{\prime} \epsilon_{i,j} \Delta x_{i,t-1} \right|
\leq 2 \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||\Phi_i - \Phi_i||^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{i=1}^{N_1} ||\epsilon_{i,j} \Delta x_{i,t-1}||^2 \right)^{1/2}
= O_p \left( \frac{1}{\sqrt{T}} \right). \tag{A14}
\]

As the above calculations make clear, \( \bar{z}_{01,1}, \sum_{i=1}^{N_1} \epsilon_{i,j} d_i' HF_i/N_1 \) in \( \bar{z}_{41,1} \) and \( \bar{z}_{61,1} \) are the leading terms in \( \bar{z}_{1,1} \). Hence, by direct substitution of (A8)–(A12) into the definition of \( \bar{z}_{1,1} \),
\[
\bar{z}_{1,1} = \bar{z}_{01,1} + \bar{z}_{41,1} + \bar{z}_{61,1} + O_p \left( \frac{1}{C_{NT}^2} \right), \tag{A15}
\]
where the order of the remainder follows from the fact that
\[
O_p \left( \frac{1}{C_{NT}^2} \right) + O_p \left( \frac{1}{\sqrt{N_1 C_{1NT}}} \right) + O_p \left( \frac{1}{\sqrt{T C_{NT}}} \right) = O_p \left( \frac{1}{C_{NT}^2} \right).
\]
Because \( \bar{\varepsilon}_{0,l} + \bar{\varepsilon}_{41,l} + \bar{\varepsilon}_{61,l} = O_p(1/C_{NT}) \), the remainder term in (A7) is \( O_p(1/C_{NT}^3) \), and therefore
\[
\bar{w}_{1,l} = w_{1,l} - \frac{1}{2} w^3_{1,l} \bar{\varepsilon}_{1,l} + \frac{3}{4} w^5_{1,l} \bar{\varepsilon}_{2,l} + O_p \left( \frac{1}{C_{NT}^3} \right).
\] (A16)

The result in (A16) suggests that \( \Delta \hat{R}_{i,l} \) can be expanded as
\[
\Delta \hat{R}_{i,l} = \bar{w}_{1,l} \hat{e}_{i,l} = w_{1,l} (\epsilon_{i,l} + (\rho_i - 1)\rho_i w_{i,l} - (\Phi_i - \Phi_i)')\Delta x_{i,l-1} - \frac{1}{2} w^3_{1,l} \bar{\varepsilon}_{1,l} \Delta R_{wi,l} - \frac{3}{4} w^5_{1,l} \bar{\varepsilon}_{2,l} \Delta R_{wi,l} + O_p \left( \frac{1}{C_{NT}^3} \right),
\] (A17)

where \( \Delta R_{wi,l} = \epsilon_{i,l} + (\rho_i - 1)\rho_i w_{i,l} \). Consider the second term on the right-hand side, \( w_{1,l} (\epsilon_{i,l} + (\Phi_i - \Phi_i)')\Delta x_{i,l-1} \). By using \( a_{i,l} = \Theta_i H^{-1} \epsilon_i - d_i \hat{p}_i = \Theta_i H^{-1} \epsilon_i + O_p(1/C_{NT}) \) and \( ||(\Phi_i - \Phi_i)\Delta x_{i,l-1}|| \leq ||\Phi_i - \Phi_i|| ||\Delta x_{i,l-1}|| = O_p(1/\sqrt{T}) = O_p(1/C_{2NT}) \), it is clear that
\[
w_{1,l} (\epsilon_{i,l} + (\Phi_i - \Phi_i)')\Delta x_{i,l-1} = \Theta_i H^{-1} w_{1,l} \epsilon_i + O_p \left( \frac{1}{C_{2NT}} \right).
\]

The order of \( \bar{\varepsilon}_{1,l} \Delta R_{wi,l} \) is equal to that of \( \bar{\varepsilon}_{1,l} \), and therefore
\[
\bar{\varepsilon}_{1,l} \Delta R_{wi,l} = (\bar{\varepsilon}_{0,l} + \bar{\varepsilon}_{41,l} + \bar{\varepsilon}_{61,l}) \Delta R_{wi,l} + O_p \left( \frac{1}{C_{NT}^2} \right).
\]

Similarly, since \( \bar{\varepsilon}_{1,l} \) and \( a_{i,l} \) are both \( O_p(1/C_{NT}) \),
\[
||\bar{\varepsilon}_{1,l} (a_{i,l} - (\Phi_i - \Phi_i)\Delta x_{i,l-1})|| \leq ||\bar{\varepsilon}_{1,l}|| ||a_{i,l} - (\Phi_i - \Phi_i)|| ||\Delta x_{i,l-1}|| = O_p \left( \frac{1}{C_{NT}^3} \right) + O_p \left( \frac{1}{\sqrt{T}C_{NT}} \right),
\]

which is clearly \( O_p(1/C_{2NT}) \). The other terms are all of higher order than this. It follows that
\[
\Delta \hat{R}_{i,l} = \Delta R_{wi,l} + \Theta_i H^{-1} w_{1,l} \epsilon_i - \frac{1}{2} w^3_{1,l} (\bar{\varepsilon}_{0,l} + \bar{\varepsilon}_{41,l} + \bar{\varepsilon}_{61,l}) \Delta R_{wi,l} + O_p \left( \frac{1}{C_{2NT}} \right)
\]
\[
= \Delta R_{wi,l} + \Theta_i H^{-1} w_{1,l} \epsilon_i - \frac{1}{2} w^3_{1,l} (\bar{\varepsilon}_{0,l} + \bar{\varepsilon}_{41,l} + \bar{\varepsilon}_{61,l}) \epsilon_i + \Theta_i H^{-1} \epsilon_i + O_p \left( \frac{1}{C_{2NT}} \right)
\]
\[
= \Delta R_{wi,l} + \Theta_i H^{-1} w_{1,l} \epsilon_i + \Theta_i H^{-1} \epsilon_i + O_p \left( \frac{1}{C_{2NT}} \right),
\] (A18)

with an obvious definition of \( g_{i,l} \), and where the second equality uses the approximation \( \Delta R_{wi,l} = \epsilon_{i,l} + (\rho_i - 1)w_{1,l} \rho_i - \epsilon_{i,l} + O_p(1) \), 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 (A16),
\[
\frac{1}{\sqrt{T}} \hat{R}_{wi,t} = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} \Delta \hat{R}_{wi,s} = \frac{1}{\sqrt{T}} R_{wi,t} + \sum_{s=p_i+2}^{t} \frac{1}{\sqrt{T}} \sum_{i} w_{1,s}(a_{i,s} - (\Phi_i - \Phi_t)') \Delta x_{i,s-1} - \frac{1}{2\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}^3 \bar{c}_{1,s} \Delta R_{wi,s} - \frac{1}{2\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}^4 \bar{c}_{1,s}(a_{i,s} - (\Phi_i - \Phi_t)') \Delta x_{i,s-1} + \frac{3}{4\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}^5 \bar{c}_{1,s}^2 \Delta R_{wi,s} + \frac{3}{4\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}^6 \bar{c}_{1,s}^3(a_{i,s} - (\Phi_i - \Phi_t)') \Delta x_{i,s-1} + O_p\left(\frac{\sqrt{T}}{C_{NT}}\right),
\]
(A19)
where again all terms that are \(O_p(1/C_{2NT})\) can be treated as negligible.

Consider \(\sum_{s=p_i+2}^{t} w_{1,s}(a_{i,s} - (\Phi_i - \Phi_t)') \Delta x_{i,s-1} / \sqrt{T}\). Via \(\hat{f}_s = HF_s + v_s\), and then leading term approximation, \(\sum_{s=p_i+2}^{t} w_{1,s} \hat{f}_s / \sqrt{T} = \sum_{s=p_i+2}^{t} w_{1,s} HF_s / \sqrt{T} + o_p(1) = O_p(1)\), implying
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s} a_{i,s} = \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s} v_s - d_i \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s} \hat{f}_s = \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s} v_s + O_p\left(\frac{1}{C_{2NT}}\right),
\]
which, together with \((\Phi_i - \Phi_t)') \sum_{s=p_i+2}^{t} w_{1,s} \Delta x_{i,s-1} / \sqrt{T} = O_p(1 / \sqrt{T})\), yields
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}(a_{i,s} - (\Phi_i - \Phi_t)') \Delta x_{i,s-1} = \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s} v_s + O_p\left(\frac{1}{C_{2NT}}\right). \quad (A20)
\]
Next, consider
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}^3 \bar{c}_{1,s} \Delta R_{wi,s} = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{1,s}^3 (\bar{c}_{01,s} + \cdots + \bar{c}_{91,s}) \Delta R_{wi,s},
\]
(A21)
where \(\sum_{s=p_i+2}^{t} w_{1,s}^3 \bar{c}_{01,s} \Delta R_{wi,s} / \sqrt{T}\) is again not of sufficiently low order to be ignored. Let us therefore consider \(\sum_{s=p_i+2}^{t} w_{1,s}^3 \bar{c}_{61,s} \Delta R_{wi,s} / \sqrt{T}\). Because \(\epsilon_{i,t}\) is serially uncorrelated and independent of \(F_t\), we have \(E(\epsilon_{i,s} \Delta x_{k,s-1}) = 0\). Moreover, \(\sum_{s=p_i+2}^{t} w_{1,s}^2 \epsilon_{i,s} \epsilon_{i,s} \Delta x_{k,s-1} / \sqrt{T} = \)
\(O_p(1)\), implying
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^3 \varepsilon_{61,s} \varepsilon_{i,s}
\]
\[
= \frac{1}{N_1} \sum_{k=1}^{N_1} (\Phi_k - \Phi_k)' \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^2 \varepsilon_{k,s} \Delta x_{k,s-1} \varepsilon_{i,s}
\]
\[
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \|\Phi_k - \Phi_k\|^2 \right)^{1/2} \left( \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^2 \varepsilon_{k,s} \Delta x_{k,s-1} \right)^{1/2}
\]
\[
= O_p\left(\frac{1}{\sqrt{T}}\right).
\]
Similarly, since \(\sum_{s=p_i+2}^t w_{1,s}^3 \varepsilon_{k,s} \Delta x_{k,s-1} R_{i,s-1} / T^{3/2} = O_p(1)\), we can show
\[
(\rho_i - 1) \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^4 \varepsilon_{61,s} R_{i,s-1} = \frac{1}{N_1} \sum_{k=1}^{N_1} (\Phi_k - \Phi_k)' \frac{1}{T^{3/2}} \sum_{s=p_i+2}^t w_{1,s}^3 \varepsilon_{k,s} \Delta x_{k,s-1} R_{i,s-1}
\]
\[
= O_p\left(\frac{1}{\sqrt{T}}\right),
\]
and therefore
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^3 \varepsilon_{61,s} \Delta R_{w_{i,s}} = O_p\left(\frac{1}{\sqrt{T}}\right),
\]
which is \(O_p(1/C_{2NT})\). The same steps can be used to show that
\[
\frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 \varepsilon_{k,s} d_k' H F_s \Delta R_{w_{i,s}} = \frac{1}{N_1} \sum_{k=1}^{N_1} d_k' \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t H F_s w_{1,s}^2 \varepsilon_{k,s} \varepsilon_{i,s}
\]
\[
+ \frac{c_i}{N_1} \sum_{k=1}^{N_1} d_k' \frac{1}{T^{3/2}} \sum_{s=p_i+2}^t \varepsilon_{k,s} w_{1,s}^3 H F_s R_{i,s-1}
\]
\[
= O_p\left(\frac{1}{C_{2NT}}\right).
\]
Hence, since \(-2 \sum_{k=1}^{N_1} \varepsilon_{k,s} d_k' H F_s / N_1\) is the leading term in \(\tilde{\tau}_{4t,s}\)
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^3 \varepsilon_{4t,s} \Delta R_{w_{i,s}} = -\frac{2}{\sqrt{T}N_1} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 \varepsilon_{k,s} d_k' H F_s \Delta R_{w_{i,s}} + o_p(1) = O_p\left(\frac{1}{C_{2NT}}\right).
\]
Thus, while non-negligible in \(\Delta \hat{R}_{w_{i,t}}\), \(\varepsilon_{4t,s}\) and \(\varepsilon_{61,s}\) are in fact negligible in \(\hat{R}_{w_{i,t-1}} / \sqrt{T}\). We now verify that the remaining terms in (A21) are negligible too. In so doing, note that, as in the above, the effect of \((\rho_i - 1)w_{1,s} R_{i,s-1}\) in \(\Delta R_{w_{i,s}}\) will generally be dominated by that of \(\varepsilon_{i,s}\). Hence, unless otherwise stated, in what follows we are going to work with the approximation \(\Delta R_{w_{i,s}} = \varepsilon_{i,s} + o_p(1)\).
Both $R_{ij}/\sqrt{T}$ and $R_{wi,i}/\sqrt{T}$ satisfy an invariance principle. By using this and the fact that $(\rho_i - 1) = O_p(1/N^cT)$,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s}^3 \tau_{i1,s} \Delta R_{wi,s} = \frac{1}{\sqrt{T}N_1} \sum_{k=1}^{N_1} \sum_{s=p_i+2}^t w_{i,s}^3 (\rho_k - 1)^2 R_{k,s-1}^2 \Delta R_{wi,s} = \frac{c_i}{T^{5/2}N_{2i}^2} \sum_{k=1}^{N_1} \sum_{s=p_i+2}^t w_{i,s}^3 R_{k,s-1}^2 \Delta R_{wi,s} = O_p\left(\frac{1}{N^2\sqrt{T}}\right).
\]

It is also not difficult to show that

\[
\left| \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s}^3 \tau_{31,s} \Delta R_{wi,s} \right|
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s}^3 \hat{\epsilon}_{i,s} \Delta x_{k,s-1} \Delta x_{k,s-1}' (\Phi_k - \Phi_k) \Delta R_{wi,s} \right| \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||\Phi_k - \Phi_k||^4 \right)^{1/2}
\leq O_p\left(\frac{1}{T}\right),
\]

and similarly,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s}^3 \tau_{51,s} \Delta R_{wi,s} = \frac{2}{\sqrt{T}N_1} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{i,s}^3 (\rho_k - 1) \hat{a}_{k,s} R_{k,s-1} \Delta R_{wi,s} = O_p\left(\frac{1}{N^{1/2+1/2\sqrt{T}}}\right),
\]

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s}^3 \tau_{91,s} \Delta R_{wi,s} = -\frac{2}{\sqrt{T}N_1} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{i,s}^3 (\rho_k - 1) (\Phi_k - \Phi_k) R_{k,s-1} \Delta x_{k,s-1} \Delta R_{wi,s} = O_p\left(\frac{1}{N^cT}\right),
\]

which are all $O_p(1/CNT)$.

The terms involving $a_{ij}$ can be evaluated in a similar fashion. In particular, we have

\[
\left| \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{i,s}^3 \tau_{71,s} \Delta R_{wi,s} \right|
\leq 2 \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| \frac{1}{N_1^{1/2}} \sum_{s=p_i+2}^t w_{i,s}^3 R_{k,s-1} \Delta R_{wi,s} \right| \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left\| \sum_{s=p_i+2}^t a_{k,s} \right\| \right)^{1/2}
\leq O_p\left(\frac{1}{N^c\sqrt{T}C_{NT}}\right),
\]

35
where we have made use of the fact that $\sum_{s=p_t+2}^{t} w_{1,s}^3 R_{k,s-1} \Delta R_{w_i,s} / T = O_p(1)$ and the results of Bai and Ng (2004, page 1158), from which we deduce that $\sum_{s=p_t+2}^{t} a_{k,s} / \sqrt{T} = O_p(1/C_{NT})$.

Clearly, $O_p(1/N^{\epsilon} \sqrt{T} C_{NT}) = O_p(1/C_{2NT})$.

Moreover, since $c_{2,t} = a_{i,t}^2 = (\Theta'_t H^{-1} v_t)^2 - 2\Theta'_t H^{-1} v_t d'_t \hat{F}_t + (d'_t \hat{F}_t)^2$,

$$\frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^3 c_{21,s} \Delta R_{w_i,s}$$

$$= \frac{1}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 a_{k,s}^2 \Delta R_{w_i,s}$$

$$= \frac{1}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\Theta'_k H^{-1} v_s)^2 \Delta R_{w_i,s} - \frac{2}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 \Theta'_k H^{-1} v_s d'_k \hat{F}_s \Delta R_{w_i,s}$$

$$+ \frac{1}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (d'_k \hat{F}_s)^2 \Delta R_{w_i,s}.$$ 

Now, since $\Theta'_k H^{-1} v_s$ and $d'_k \hat{F}_s$ are just scalars, $\Theta'_k H^{-1} v_s d'_k \hat{F}_s = \hat{F}_s v'_s (H^{-1})' \Theta_k d'_k$. Hence, since by $||a + b||^2 \leq ||a||^2 + ||b||^2$, $||\sum_{s=p_t+2}^{t} w_{1,s}^3 \hat{F}_s \Delta R_{w_i,s} v'_s||^2 \leq \sum_{s=p_t+2}^{t} w_{1,s}^3 (\Delta R_{w_i,s} v'_s)^2 ||\hat{F}_s||^2 ||v_s||^2$, suggesting that

$$\left| \frac{1}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 \Theta'_k H^{-1} v_s d'_k \hat{F}_s \Delta R_{w_i,s} \right|$$

$$\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \sum_{s=p_t+2}^{t} w_{1,s}^3 (\Delta R_{w_i,s})^2 ||\hat{F}_s||^2 ||v_s||^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||\Theta_k||^2 ||d_k||^2 \right)^{1/2} \left| H^{-1} \right|$$

$$= O_p\left( \frac{1}{C_{1NT} C_{2NT}} \right).$$

Also, by the cyclical property of the trace, $tr((d'_k \hat{F}_s)^2) = tr(d'_k \hat{F}_s d'_k) = tr(d_k d'_k \hat{F}_s \hat{F}_s')$. Therefore, since $\Delta R_{w_i,t} = \epsilon_{i,t} + o_p(1)$, where $\epsilon_{i,t}$ is asymptotically uncorrelated with $\hat{F}_t$,

$$\frac{1}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (d'_k \hat{F}_s)^2 \Delta R_{w_i,s} = tr \left( \frac{1}{N_1} \sum_{k=1}^{N_1} d_k d'_k \frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^3 \hat{F}_s \hat{F}_s' \epsilon_{i,t} \right) = O_p\left( \frac{1}{C_{2NT}} \right),$$

suggesting

$$\frac{1}{\sqrt{T}} \sum_{s=p_t+2}^{t} w_{1,s}^3 c_{21,s} \Delta R_{w_i,s} = \frac{1}{\sqrt{T} N_1} \sum_{s=p_t+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\Theta'_k H^{-1} v_s)^2 \Delta R_{w_i,s} + O_p\left( \frac{1}{C_{NT}} \right).$$

As for the remaining term, according to Bai (2003, page 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 $\epsilon_{i,s}$ is cross-section

36
independent, we obtain

$$
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} w_{1,s}^3 v_s v_s' \epsilon_{i,s}
$$

$$
= H' \Lambda^{-1} \frac{1}{\sqrt{T} N^2} \sum_{s=p_1+2}^{t} \sum_{k=1}^{N} \sum_{n=1}^{N} \Theta_k w_{1,s}^3 \epsilon_{k,s} \epsilon_{n,s} \epsilon_{i,s} \Theta_n' (\Lambda^{-1})' H + o_p(1)
$$

$$
= H' \Lambda^{-1} \frac{1}{\sqrt{T} N^2} \sum_{s=p_1+2}^{t} \sum_{k=1}^{N} \Theta_k w_{1,s}^3 \epsilon_{k,s}^2 \epsilon_{i,s} \Theta_k' (\Lambda^{-1})' H + o_p(1)
$$

$$
= H' \Lambda^{-1} \frac{1}{\sqrt{T} N^2} \sum_{s=p_1+2}^{t} \Theta_k w_{1,s}^5 \epsilon_{i,s}^3 \Theta_k' (\Lambda^{-1})' H + o_p(1)
$$

$$
= H' \Lambda^{-1} \frac{1}{\sqrt{T} N^2} \sum_{s=p_1+2}^{t} \Theta_k w_{1,s}^5 \epsilon_{i,s}^3 \Theta_k' (\Lambda^{-1})' H + o_p(1)
$$

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

where \( \Lambda_s = \lim N \to \infty \sum_{k=1}^{N} \sigma_{k,s}^2 \Theta_k \Theta_k' / N \) such that \( \Lambda_{[rT]} \to \Lambda(r) > 0 \) as \( T \to \infty \). Therefore, since \( tr((\Theta_k' H^{-1} v_s)^2) = tr(v_s v_s' (H^{-1})' \Theta_k \Theta_k' H^{-1}) \) and \( \Delta R_{wi,s} = \epsilon_{i,s} + o_p(1) \), we can show that

$$
\frac{1}{\sqrt{T} N_1} \sum_{s=p_1+2}^{t} \sum_{k=1}^{N_1} w_{1,s}^3 (\Theta_k' H^{-1} v_s)^2 \Delta R_{wi,s}
$$

$$
= tr\left(\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} w_{1,s}^3 v_s v_s' \epsilon_{i,s} \frac{1}{N_1} \sum_{k=1}^{N_1} (H^{-1})' \Theta_k \Theta_k' H^{-1}\right) + o_p(1) = O_p\left(\frac{\sqrt{T}}{C_{NT}^3}\right),
$$

from which it follows that

$$
\frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{t} w_{1,s}^3 v_{21,s} \Delta R_{wi,s} = O_p\left(\frac{\sqrt{T}}{C_{NT}}\right).
$$
As for the term involving $\zeta_{81,s}$,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^3 \zeta_{81,s} \Delta R_{wi,s} = -\frac{2}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 a_{k,s} (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \Delta R_{wi,s}
\]

\[
= -\frac{2}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 \Theta_k' H^{-1} v_s (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \epsilon_{i,s}
\]

\[
= \frac{2}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 \Theta_k' H^{-1} v_s (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \epsilon_{i,s} + o_p(1)
\]

\[
= -\frac{2}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 \Theta_k' H^{-1} v_s (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \epsilon_{i,s} + O_p\left(\frac{1}{\sqrt{TC_{2NT}}}\right),
\]

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

\[
\left| \frac{1}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 d_k' \hat{F}_s (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \epsilon_{i,s} \right| \leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||d_k||^2 ||\Phi_k - \Phi_k||^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^t w_{1,s}^3 \hat{F}_s \Delta x_{k,s-1} \epsilon_{i,s} \right| \right)^{1/2}
\]

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

The remaining term can be written as

\[
\frac{1}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 \Theta_k' H^{-1} v_s (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \epsilon_{i,s}
\]

\[
= tr\left( \frac{1}{\sqrt{TN_1}} \sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 v_s (\Delta x_{k,s-1})' (\Phi_k - \Phi_k) \Theta_k' H^{-1} \right).
\]

If $k = i$,

\[
\frac{1}{\sqrt{TN}} \sum_{s=p_i+2}^t \sum_{k=1}^{N} w_{1,s}^3 \epsilon_{i,s} (\Delta x_{k,s-1})' = \frac{1}{\sqrt{TN}} \sum_{s=p_i+2}^t w_{1,s}^3 \epsilon_{i,s} (\Delta x_{i,s-1})',
\]

which is $O_p(1/N)$, suggesting that $\sum_{s=p_i+2}^t \sum_{k=1}^{N_1} w_{1,s}^3 v_s \epsilon_{i,s} (\Delta x_{k,s-1})' / \sqrt{T}$ is of the same order.
If \( k \neq i \), then the same quantity is \( O_p(1/\sqrt{N}) \). It follows that
\[
\left| \frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^3 \Theta_k H^{-1} v_s (\Phi_k - \Phi_k)' \Delta x_{k,s-1} \epsilon_{i,s} \right| \leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^3 v_s \epsilon_{i,s} (\Delta x_{k,s-1})' \right|^2 \right)^{1/2} \times \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||\Theta_k||^2 ||\Phi_k||^2 \right)^{1/2} \|H^{-1}\| = O_p\left( \frac{1}{\sqrt{NT}} \right),
\]
and therefore
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^3 \epsilon_{61,s} \Delta R_{wi,s} = O_p\left( \frac{1}{\sqrt{NT}} \right).
\]
Thus, by combining the results,
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^3 \epsilon_{1,s} \Delta R_{wi,s} = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^3 (\epsilon_{11,s} + \ldots + \epsilon_{61,s}) \Delta R_{wi,s} = O_p\left( \frac{\sqrt{T}}{C_{NT}} \right) + O_p\left( \frac{1}{C_{2NT}} \right).
\]
Moreover, by using the same steps as in the above, the fourth term on the right-hand side of \( (A19) \) is
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^3 \epsilon_{1,s} (a_{i,s} - (\Phi_i - \Phi_i)' \Delta x_{i,s-1}) = O_p\left( \frac{1}{C_{NT}} \right),
\]
where \( O_p(1/C_{2NT}) > O_p(1/C_{NT}). \) As for the fifth term, \( \sum_{s=p_i+2}^{t} w_{i,s}^5 \epsilon_{01,s}^2 \Delta R_{wi,s} / \sqrt{T}, \) by using \( \epsilon_{i,s} = \epsilon_{01,s} + \epsilon_{41,s} + \epsilon_{61,s} + o_p(1) \) and \( \Delta R_{wi,s} = \epsilon_{i,s} + o_p(1), \) we get
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \epsilon_{01,s}^2 \Delta R_{wi,s} = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 (\epsilon_{01,s} + \epsilon_{41,s} + \epsilon_{61,s})^2 \epsilon_{i,s} + o_p(1).
\]
Consider \( \sum_{s=p_i+2}^{t} w_{i,s}^5 \epsilon_{01,s}^2 \epsilon_{i,s} / \sqrt{T}. \) From \( \epsilon_{01,s} = \sum_{i=1}^{N_1} \sigma_{1,i}^2 (\epsilon_{i,s}^2 - 1)^2 / N_1^2 + o_p(1) = O_p(1/N), \) we obtain
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \epsilon_{01,s}^2 \epsilon_{i,s} = \frac{1}{\sqrt{T}N_1^2} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^5 (\epsilon_{k,s}^2 - 1)^2 \epsilon_{i,s} + o_p(1) = O_p\left( \frac{1}{N} \right).
\]
Also,
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \epsilon_{01,s} (\epsilon_{61,s} + \epsilon_{41,s}) \epsilon_{i,s} = -\frac{2}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^5 \epsilon_{01,s} [(\Phi_k - \Phi_k)' \Delta x_{k,s-1} + d_k' HF_k] \epsilon_{k,s} \epsilon_{i,s} + O_p\left( \frac{1}{C_{2NT}} \right).
\]
The order of the remainder is not the sharpest possible but it is enough to show that it is \( O_p(1/N^k) \). Consider the first term on the right-hand side. By cross-section independence,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \tau_{01,s} \epsilon_{k,s} \epsilon_{i,s} \Delta x_{k,s-1} = \frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{u=1}^{N_1} w_{i,s}^2 (\epsilon_{u,s}^2 - 1) \epsilon_{i,s} \epsilon_{k,s} \Delta x_{k,s-1} = O_p \left( \frac{1}{\sqrt{N}} \right),
\]

which in turn implies

\[
\left| \frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^5 \tau_{01,s} \epsilon_{k,s} \epsilon_{i,s} (\Delta x_{k,s-1})' (\Phi_k - \Phi_k) \right| 
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} \left| \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \tau_{01,s} \epsilon_{k,s} \epsilon_{i,s} (\Delta x_{k,s-1}) \right|^2 \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||\Phi_k - \Phi_k||^2 \right)^{1/2} = O_p \left( \frac{1}{\sqrt{NT}} \right).
\]

The same steps can be used to show that

\[
\frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^5 \tau_{01,s} \epsilon_{k,s} d_k' HF \epsilon_{i,s} = O_p \left( \frac{1}{\sqrt{NC_2NT}} \right),
\]

and therefore

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \tau_{01,s} \tau_{41,s} \epsilon_{i,s} = O_p \left( \frac{1}{C_2NT} \right).
\]

Also,

\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} w_{i,s}^5 \tau_{61,s} \epsilon_{i,s} = \frac{4}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^3 ((\Phi_k - \Phi_k)' \Delta x_{k,s-1})^2 \epsilon_{k,s}^2 \epsilon_{i,s} = O_p \left( \frac{1}{\sqrt{T}} \right),
\]

which is clearly \( O_p(1/C_2NT) \). As for \( \sum_{s=p_i+2}^{t} w_{i,s}^5 \tau_{41,s} \epsilon_{i,s} / \sqrt{T} \), since \( \epsilon_{i,s} \) is independent across \( i \) and also of \( F_i \), \( \sum_{s=p_i+2}^{t} F_i' d_k' \epsilon_{i,s}^2 / \sqrt{T} = O_p(1) \). Hence, since \( tr((d_k' F_i)^2) = tr(d_k' d_k' F_i F_i') \),

we get

\[
\frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} w_{i,s}^3 \epsilon_{i,s} \epsilon_{k,s}^2 (d_k' HF_i)^2
\]

\[
= tr \left( \frac{1}{\sqrt{T}N_1} \sum_{s=p_i+2}^{t} \sum_{k=1}^{N_1} d_k d_k' F_i F_i' w_{i,s}^3 \epsilon_{i,s} \epsilon_{k,s}^2 \right)
\]

\[
\leq \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||d_k||^4 \right)^{1/2} \left( \frac{1}{N_1} \sum_{k=1}^{N_1} ||F_i F_i' w_{i,s}^3 \epsilon_{i,s} \epsilon_{k,s}^2||^2 \right)^{1/2} = O_p \left( \frac{1}{C_2^{2NT}} \right),
\]

40
and therefore
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^T w_{i,s}^5 \epsilon_{i,s}^2 = \frac{4}{\sqrt{T}N_1} \sum_{s=p_i+2}^T \sum_{k=1}^{N_1} w_{i,s}^3 (d_k^I HF_s)^2 \epsilon_{i,s}^2 = O_p \left( \frac{1}{C^{2NT}} \right).
\]

Finally, consider \( \sum_{s=p_i+2}^T w_{i,s}^5 \epsilon_{i,s}^2 / \sqrt{T} \). Again, since \( \epsilon_{i,s} \) is independent across \( i \) and \( s \), we have, by a simple modification of Lemma 2 in Bai and Ng (2004),
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^T w_{i,s}^5 \epsilon_{i,s}^2 = O_p(1),
\]
from which we obtain
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^T w_{i,s}^5 \epsilon_{i,s}^2 = O_p \left( \frac{1}{\sqrt{T}C^{2NT}} \right).
\]
Thus, by adding the terms,
\[
\frac{1}{\sqrt{T}} \sum_{s=p_i+2}^T w_{i,s}^5 \epsilon_{i,s}^2 \Delta R_{wi,s} = O_p \left( \frac{1}{C^{2NT}} \right).
\]

The sixth and last term on the right-hand side of (A19), \( \sum_{s=p_i+2}^T w_{i,s}^5 \epsilon_{i,s}^2 \Delta x_{i,s-1} / \sqrt{T} \) is dominated by the other remainders and can be treated as \( o_p(1/C^{2NT}) \). Therefore, by adding (A19)–(A24),
\[
\frac{1}{\sqrt{T}} \hat{R}_{wi,t} = \frac{1}{\sqrt{T}} R_{wi,t} + \frac{1}{\sqrt{T}} G_{i,t} + O_p \left( \frac{1}{C^{2NT}} \right),
\]
where
\[
G_{i,t} = \Theta_i' H^{-1} \sum_{s=p_i+2}^T w_{i,s} v_s - \frac{1}{2} \sum_{s=p_i+2}^T w_{i,s} \epsilon_{01,s} \epsilon_{i,s}.
\]
This establishes the second result, and hence the proof of the lemma is complete. \( \blacksquare \)

**Proof of Theorem 1.**

Consider again the group consisting of the first \( N_1 \) observations. By Lemma A.1, with \( \kappa = 0 \),
\[
\Delta \hat{R}_{wi,t} = \Delta R_{wi,t} + G_{i,t} + O_p \left( \frac{1}{C^{2NT}} \right) = \Delta R_{wi,t} + O_p \left( \frac{1}{C^{2NT}} \right).
\]

As for \( \hat{R}_{wi,t} / \sqrt{T} \), 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, \( \sum_{s=p_{t+2}}^{t} w_{m,s} t_{0,m,s} \epsilon_{i,s} / \sqrt{T} = O_p(1/\sqrt{N}) \),

\[
\frac{1}{\sqrt{T}} \hat{R}_{w,i,t} = \frac{1}{\sqrt{T}} (R_{w,i,t} + G_{i,t}) + O_p \left( \frac{1}{C2NT} \right) = \frac{1}{\sqrt{T}} R_{w,i,t} + O_p \left( \frac{1}{CNT} \right). \tag{A27}
\]

Consider \( R_{i,t} / \sqrt{T} \). Since \( R_{i,t} = \rho_i R_{i,t-1} + \epsilon_{i,t} \), where \( \rho_i = \exp(c_i/T) \) and \( \epsilon_{i,t} = \Theta_i F_t + \epsilon_{i,t} \),

by recursive substitution,

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

\[
= \Theta_i' \frac{1}{\sqrt{T}} \sum_{k=p_{t+2}}^{P} \exp \left( \frac{c_i}{T} \right)^{t-k} F_k + \frac{1}{\sqrt{T}} \sum_{k=p_{t+2}}^{P} \exp \left( \frac{c_i}{T} \right)^{t-k} \epsilon_{i,k} + o_p(1), \tag{A28}
\]

from which we obtain

\[
\frac{1}{\sqrt{T}} R_{i,t} \to_d \Theta_i' B_{F,i}(s) + B_{\epsilon,i}(s) = B_{R,i}(s) \tag{A29}
\]

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

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

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

with \( W_{\epsilon,i}(s) \) and \( W_{F}(s) \) being two independent Brownian motions. Hence, since \( R_{w,i,t} = \sum_{s=p_{t+2}}^{t} \Delta R_{w,i,k} \), we obtain

\[
\frac{1}{\sqrt{T}} \hat{R}_{w,i,t} = \frac{1}{\sqrt{T}} \sum_{k=p_{t+2}}^{t} \epsilon_{i,k} + \frac{c_i}{T^{3/2}} \sum_{k=p_{t+2}}^{t} w_{1,k} R_{i,k-1} + O_p \left( \frac{1}{C_{NT}} \right)
\]

\[
\to_d W_{\epsilon,i}(s) + c_i \int_{r=0}^{s} w_1(r) B_{R,i}(r) dr = V_{R,i}(s) \tag{A30}
\]

as \( N, T \to \infty \). This result, together with the continuous mapping theorem, imply

\[
\frac{1}{T} \sum_{t=p_{t+2}}^{T} \hat{R}_{w,i,t-1} \Delta R_{w,i,t} = \frac{1}{T} \sum_{t=p_{t+2}}^{T} R_{w,i,t-1} \Delta R_{w,i,t} + O_p \left( \frac{1}{C_{NT}} \right) \to_d \int_{s=0}^{1} V_{R,i}(s) dV_{R,i}(s). \tag{A31}
\]

The proof is completed by noting that

\[
\frac{1}{T^2} \sum_{t=p_{t+2}}^{T} \hat{R}_{w,i,t-1}^2 \to_d \frac{1}{T^2} \sum_{t=p_{t+2}}^{T} R_{w,i,t-1}^2 + O_p \left( \frac{1}{C_{NT}} \right) \to_d \int_{s=0}^{1} V_{R,i}(s)^2 ds \tag{A32}
\]

as \( N, T \to \infty \).

\[ \blacksquare \]

**Proof of Theorem 2.**
Consider the numerator of \( \tau \). By definition, \( \hat{R}_{wi,t} = \sum_{s=p_i+2}^T \Delta \hat{R}_{wi,s} \) or \( \hat{R}_{wi,t} = \hat{R}_{wi,t-1} + \Delta \hat{R}_{wi,t} \), showing that \( \hat{R}_{wi,t}^2 = (\hat{R}_{wi,t-1} + \Delta \hat{R}_{wi,t})^2 = \hat{R}_{wi,t-1}^2 + 2\hat{R}_{wi,t-1}\Delta \hat{R}_{wi,t} + (\Delta \hat{R}_{wi,t})^2 \). It follows that

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

\[
= \frac{1}{2} (\hat{R}_{wi,T}^2 - \hat{R}_{wi,p+1}^2) - \frac{1}{2} \sum_{t=p_i+2}^T (\Delta \hat{R}_{wi,t})^2, \tag{A33}
\]

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

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p_i+2}^T \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t} = \frac{1}{\sqrt{NT}} \sum_{m=1}^{M} \sum_{n=1}^{N_m} \sum_{t=p_i+2}^T \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t}.
\]

Hence, focusing of the first group comprising the first \( N_1 \) units,

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_i+2}^T \hat{R}_{wi,t-1} \Delta \hat{R}_{wi,t}
= \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N_1} (\hat{R}_{wi,t}^2 - \hat{R}_{wi,p+1}^2) - \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_i+2}^T (\Delta \hat{R}_{wi,t})^2. \tag{A34}
\]

It is trivial to show that \( \sum_{i=1}^{N_1} \hat{R}_{wi,p+1}^2 / \sqrt{NT} \) and \( \sum_{i=1}^{N_1} R_{wi,p+1}^2 / \sqrt{NT} \) are \( O_p(\sqrt{N}/T) \). Let us therefore consider \( \hat{R}_{wi,T}^2 \). Clearly, by Lemma A.1,

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \hat{R}_{wi,T}^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} (\hat{R}_{wi,T}^2 + 2R_{wi,T}G_{i,T} + G_{i,T}^2) + O_p\left(\frac{\sqrt{N}}{C_2NT}\right),
\]

where

\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} R_{wi,T}G_{i,T} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} R_{wi,T} \Theta_i H^{-1} \sum_{s=p_i+2}^T w_{i,s} v_s - \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_i+2}^T w_{i,s} \varepsilon_{01,s} \varepsilon_{i,s}.
\]

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

\[
\frac{1}{\sqrt{TN}} \sum_{i=1}^{N_1} R_{wi,T} \Theta_i' = \frac{1}{\sqrt{TN}} \sum_{i=1}^{N_1} \sum_{s=p_i+2}^T \varepsilon_{i,s} \Theta_i + \frac{1}{T^{3/2}N} \sum_{i=1}^{N_1} \sum_{s=p_i+2}^T c_i \Theta_i' w_{i,s} R_{i,-1} = O_p(1).
\]

Moreover, as in the proof of Theorem 1, \( \sum_{i=1}^{N_1} w_{i,s} v_s / \sqrt{T} = O_p(1/\sqrt{N}) + O_p(1/T^{3/2}) = O_p(1/C_{NT}) \), and since \( \varepsilon_{01,s} = O_p(1/\sqrt{N}) \), we also have

\[
\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{s=p_i+2}^T w_{i,s} \varepsilon_{01,s} \varepsilon_{i,s} \right| \leq \left( \frac{1}{T} \sum_{s=p_i+2}^T \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \varepsilon_{i,s} \right)^2 \left( \frac{1}{T} \sum_{s=p_i+2}^T w_{i,s}^2 \varepsilon_{01,s}^2 \right)^{1/2} = O_p\left(\frac{1}{\sqrt{N}}\right).
\]
It follows that
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} R_{wi,T} G_{i,T} = O_p\left(\frac{1}{C_{NT}}\right).
\]

Also, since by the proof of Lemma A.1, \(\sum_{i=p_1+2}^{T} w_i^3 \zeta_{01,i} \epsilon_{i,t} / \sqrt{T} = O_p(1/\sqrt{N})\), suggesting
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} G_{i,T}^2 = \frac{1}{N} \sum_{i=1}^{N_1} \left( \Theta_i H^{-1} \frac{1}{\sqrt{T}} \sum_{s=p_1+2}^{T} w_{1,s} v_s - \frac{1}{2\sqrt{T}} \sum_{s=p_1+2}^{T} w_{1,s}^3 \zeta_{01,s} \epsilon_{i,s} \right)^2 = O_p\left(\frac{\sqrt{N}}{C_{2NT}}\right).
\]

It follows that
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \hat{R}_{wi,T}^2 = O_p\left(\frac{\sqrt{N}}{C_{2NT}}\right). \quad (A35)
\]

Next, consider \(\sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} (\Delta\hat{R}_{wi,t})^2 / \sqrt{NT}\) in (A33). By Lemma A.1,
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} (\Delta\hat{R}_{wi,t})^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} \left((\Delta R_{wi,t})^2 - 2\Delta R_{wi,t} g_{i,t} + \sigma_{1,t}^2\right) + O_p\left(\frac{\sqrt{N}}{C_{2NT}}\right),
\]
where
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} \Delta R_{wi,t} g_{i,t} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} \Delta R_{wi,t} \Theta_i H^{-1} w_{1,t} v_t
\]
\[
- \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} w_{1,t}^3 (\zeta_{01,t} + \zeta_{41,t} + \zeta_{61,t}) \epsilon_{i,t}.
\]

From Lemma A.1 we have that \(\sum_{i=1}^{N_1} \Delta R_{wi,t} \Theta_i \sqrt{N} = O_p(1)\) and \(||v_t|| = O_p(1/C_{1NT})\), and therefore
\[
\left| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} \Delta R_{wi,t} \Theta_i H^{-1} w_{1,t} v_t \right|
\]
\[
\leq \left( \frac{1}{T} \sum_{s=p_1+2}^{T} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \Delta R_{wi,t} \Theta_i \right|^2 \right)^{1/2} \left( \frac{1}{T} \sum_{s=p_1+2}^{T} \|w_{1,t}\|^2 \|v_t\|^2 \right)^{1/2} ||H^{-1}||
\]
\[
= O_p\left(\frac{1}{C_{1NT}}\right).
\]

Similarly, since \((\zeta_{01,t} + \zeta_{41,t} + \zeta_{61,t}) = O_p(1/C_{NT})\) and \(\sum_{i=1}^{N_1} \epsilon_{i,t} / \sqrt{N} = O_p(1)\), we can show that
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} w_{1,t}^3 (\zeta_{01,t} + \zeta_{41,t} + \zeta_{61,t}) \epsilon_{i,t} = O_p\left(\frac{1}{C_{NT}}\right),
\]
from which we obtain
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} (\Delta\hat{R}_{wi,t})^2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} (\Delta R_{wi,t})^2 + O_p\left(\frac{\sqrt{N}}{C_{2NT}}\right). \quad (A36)
\]
Thus, by adding the results,
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} \hat{R}_{wi,t} \Delta \hat{R}_{wi,t} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} R_{wi,t} \Delta R_{wi,t} + O_p \left( \frac{\sqrt{N}}{C_{2NT}} \right). 
\]

Consider \( \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} R_{wi,t} \Delta R_{wi,t} / \sqrt{NT} \). By using \( \kappa = 1/2 \), and the definition of \( \Delta R_{wi,t} \),
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} R_{wi,t} \Delta R_{wi,t} 
= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=p_i+2}^{T} R_{wi,t-1} \epsilon_{i,t}. 
\]
As for the first term, since \( \kappa = 1/2 \) in this case, \( \rho_i = \exp(c_i / \sqrt{NT}) = 1 + c_i / \sqrt{NT} + o_p(1) \), which in turn suggests
\[
E \left( \frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} R_{wi,t-1} R_{i,t-1} | c_i \right) 
= \frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} \sum_{s=p_i+2}^{t-1} E(\epsilon_{i,s} R_{i,t-1} | c_i) 
+ \frac{1}{\sqrt{NT^3}} \sum_{i=1}^{N} c_i w_{1,t} \sum_{s=p_i+2}^{t-1} w_{t,s} E(R_{i,s-1} R_{i,t-1} | c_i) + o_p(1). 
\]
Consider \( R_{i,t} / \sqrt{T} \). By substitution of \( \rho_i = 1 + c_i / \sqrt{NT} + o_p(1) \),
\[
\frac{1}{\sqrt{T}} R_{i,t} = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} \rho_i^{t-s} e_{s,t} + o_p(1) = \frac{1}{\sqrt{T}} \sum_{s=p_i+2}^{t} e_{s,t} + o_p(1). 
\]
Moreover, since \( \epsilon_{i,t} \) and \( F_t \) are independent, \( E(\epsilon_{i,s} \epsilon_{i,k}) = E(\epsilon_{i,s} (\Theta_i F_k + \epsilon_{i,k})) = \sigma_{1,k} E(\epsilon_{i,s} \epsilon_{i,k}) \), which is zero if \( s \neq k \) and \( \sigma_{1,k} \) if \( s = k \). Letting
\[
g_1(v) = w_1(v) \int_0^v \sigma_1(r) du, 
\]
these results imply
\[
\frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} \sum_{s=p_i+2}^{t-1} E(\epsilon_{i,s} R_{i,t-1}) = \frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} \sum_{s=p_i+2}^{t-1} \sum_{k=p_i+2}^{t-1} E(\epsilon_{i,s} \epsilon_{i,k}) 
= \frac{1}{T^2} \sum_{t=p_i+2}^{T} w_{1,t} \sum_{k=p_i+2}^{t} \sigma_{1,k} \rightarrow \int_0^1 g_1(v) dv 
\]
as \( T \rightarrow \infty \). But we also have \( E(\epsilon_{i,s} \epsilon_{i,k}) = \Theta_i^t E(F_k^t) \Theta_i + \sigma_{1,s} \sigma_{1,k} E(\epsilon_{i,s} \epsilon_{i,k}) \), which is zero if \( s \neq k \) and \( (\Theta_i^t \Theta_i + \sigma_{i,k}^2) \) if \( s = k \). Hence, with \( t \geq s \),
\[
E(R_{i,s} R_{i,t}) = \sum_{k=p_i+2}^{s} \sum_{n=p_i+2}^{t} E(\epsilon_{i,k} \epsilon_{i,n}) = \sum_{k=p_i+2}^{s} E(\epsilon_{i,k}^2) = \sum_{k=p_i+2}^{s} (\Theta_i^t \Theta_i + \sigma_{i,k}^2), 
\]
45
which we can use to show that
\[
\frac{1}{T^3} \sum_{t=p_1+2}^{T} \sum_{s=p_1+2}^{t-1} w_{1,s} E(R_{i,s-1}R_{i,t-1} | c_i) = \frac{1}{T^3} \sum_{t=p_1+2}^{T} \sum_{s=p_1+2}^{t-1} w_{1,s} \sum_{k=p_1+2}^{s-1} (\Theta' \Sigma \Theta + \sigma^2_i) \\
\rightarrow \int_{u=0}^{1} \int_{v=0}^{u} g_i(u,v) dv du
\]
as \( T \to \infty \), where
\[
g_i(u,v) = w_1(u)w_1(v) \int_{x=0}^{v} (\Theta' \Sigma(x) \Theta + \sigma_1(x)^2) dx,
\]
Let us also define
\[
q_{N,i}(u) = \mu c g_1(u) + \frac{(\mu^2 + \sigma^2)}{\sqrt{N}} \int_{v=0}^{u} g_i(u,v) dv,
\]
\[
\bar{q}_{N,1}(u) = \frac{1}{N_1} \sum_{i=1}^{N_1} q_{N,i}(u),
\]
with \( q_i(u) \) and \( \bar{q}_1(u) \) being the corresponding integrals after taking \( N \to \infty \). That is, \( q_i(u) = \bar{q}_1(u) = \mu c g_1(u) \). By using this and the fact that \( E(c_i^2) = \mu^2 + \sigma^2 \), we obtain
\[
E \left( \frac{1}{T} \sum_{t=p_1+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \right) \rightarrow \int_{u=0}^{1} q_i(u) du,
\]
and by application of Corollary 1 of Phillips and Moon (1999), with \( \lfloor \lambda_1 N \rfloor = N_1 \), we can further show that
\[
\frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1} \rightarrow_p \lambda_1 \int_{u=0}^{1} \bar{q}_1(u) du
\]
as \( N, T \to \infty \), where \( \rightarrow_p \) signifies convergence in probability. Thus, considering the whole cross-section,
\[
\frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=p_1+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1}
\]
\[
= \sum_{m=1}^{M+1} \frac{(\lambda_m - \lambda_{m-1})}{(N_m - N_{m-1})T^2} \sum_{i=1}^{N} \sum_{t=p_1+2}^{T} c_i w_{1,t} R_{wi,t-1} R_{i,t-1}
\]
\[
\rightarrow_p \sum_{m=1}^{M+1} (\lambda_m - \lambda_{m-1}) \int_{u=0}^{1} \bar{q}_m(u) du.
\]
Consider \( \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} R_{wi,t-1} \epsilon_{i,t} / \sqrt{NT} \). Clearly, \( E(R_{wi,i-1} \epsilon_{i,t}) = 0 \). As for the variance,
\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} R_{wi,t-1} \epsilon_{i,t} \right)^2 \right] = \frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{j=1}^{N} \sum_{t=p_1+2}^{T} \sum_{s=p_1+2}^{T} E(R_{wi,t-1} R_{wi,s-1}) E(\epsilon_{i,t} \epsilon_{j,s})
\]
\[
= \frac{1}{NT^2} \sum_{i=1}^{N_1} \sum_{t=p_1+2}^{T} E(R_{wi,t-1}^2).
\]
Consider $E(R_{wi,t-1}^2)$. Because of symmetry,
\[
\sum_{s=p_t+2}^{t-1} \sum_{k=p_t+2}^{t-1} w_{1,s}w_{1,k}E(R_{i,s-1}R_{i,k-1}|c_i) = \sum_{s=p_t+2}^{t-1} w_{1,s}E(R_{i,s-1}R_{i,k-1}|c_i) + 2 \sum_{s=p_t+2}^{t-1} w_{1,s}w_{1,k} \sum_{k=p_t+2}^{t-1} (\Theta'\Sigma_k\Theta_i + \sigma_{1,k}^2)
\]
\[
\to \int_{u=0}^{r} \left( g_i(u,u) + 2 \int_{v=0}^{u} g_i(u,v)dv \right) du
\]
as $T \to \infty$. Therefore,
\[
\frac{1}{T}E(R_{wi,t-1}^2|c_i) = \frac{1}{T} \sum_{s=p_t+2}^{t-1} \sum_{k=p_t+2}^{t-1} E(\epsilon_{i,s}\epsilon_{i,k}) + \frac{2c_i}{\sqrt{NT}^2} \sum_{s=p_t+2}^{t-1} \sum_{k=p_t+2}^{t-1} w_{1,s}E(R_{i,s-1}\epsilon_{i,k}|c_i) + \frac{c_i^2}{NT^3} \sum_{s=p_t+2}^{t-1} \sum_{k=p_t+2}^{t-1} w_{1,s}w_{1,k}E(R_{i,s-1}R_{i,k-1}|c_i)
\]
\[
\to r + \frac{2c_i}{\sqrt{N}} \int_{u=0}^{r} g_i(u)du + \frac{c_i^2}{N} \int_{u=0}^{r} \left( g_i(u,u) + 2 \int_{v=0}^{u} g_i(u,v)dv \right) du,
\]
suggesting that
\[
\frac{1}{T}E(R_{wi,t-1}^2) \to r + \frac{2\mu_c}{\sqrt{N}} \int_{u=0}^{r} g_i(u)du + \frac{(\mu_c^2 + \sigma_c^2)}{N} \int_{u=0}^{r} \left( g_i(u,u) + 2 \int_{v=0}^{u} g_i(u,v)dv \right) du = p_{N,i}(r),
\]
with an obvious definition of $p_{N,i}(r)$. The limit of (A43) is therefore given by
\[
E \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i=t-p_t+2}^{T} R_{wi,t-1}\epsilon_{i,t} \right)^2 \right] = \frac{1}{NT^2} \sum_{i=t-p_t+2}^{T} E(R_{wi,t-1}^2) \to \lambda_1 \int_{r=0}^{1} p_{N,i}(r)dr \quad (A44)
\]
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 $p_{N,i}(r),$
\[
\frac{1}{\sqrt{NT}} \sum_{i=t-p_t+2}^{T} R_{wi,t-1}\epsilon_{i,t} \to_d N \left( 0, \lambda_1 \int_{r=0}^{1} p_i(r)dr \right) \quad (A45)
\]
as $N, T \to \infty$. Note also that since the effect of the common component in $p_{N,i}(r)$ vanishes as $N \to \infty$, $R_{wi,t-1}\epsilon_{i,t}$ 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
For notational simplicity, in this proof, when context is clear, the dependence of \( \hat{\sigma}_1^2(n) \) and \( \hat{\sigma}_2^2(n) \) on \( n \) will be suppressed.

The criterion function is

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

Define

\[
\text{QML}_0(n) = \sum_{t=p_t+2}^{T} \left[ n \log(\sigma_{1,t}^2) + (N - n) \log(\sigma_{2,t}^2) \right],
\]
such that

\[ QML(n) - QML_0(N_1) \]

\[ = \sum_{t=p+1}^{T} \left[ (n \log(\hat{\sigma}_{1,t}^2(n)) - N_1 \log(\sigma_{1,t}^2)) + ((N-n) \log(\hat{\sigma}_{2,t}^2(n)) - (N-N_1) \log(\sigma_{2,t}^2)) \right] \]

\[ = \sum_{t=p+1}^{T} \left[ n \log \left( \frac{\hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} \right) + (n-N_1) \log(\sigma_{1,t}^2) + (N-n) \log \left( \frac{\hat{\sigma}_{2,t}^2(n)}{\sigma_{2,t}^2} \right) \right] - (n-N_1) \log(\sigma_{2,t}^2) \]

\[ = \sum_{t=p+1}^{T} \left[ n \log \left( \frac{\hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} \right) + (n-N_1) \log \left( \frac{\hat{\sigma}_{2,t}^2(n)}{\sigma_{2,t}^2} \right) + (n-N_1) \log \left( \frac{\sigma_{1,t}^2}{\sigma_{2,t}^2} \right) \right]. \quad (A49) \]

As in the proof of Proposition 1, we start with

\[ \hat{\sigma}_{m,t}^2 = \sigma_{m,t}^2 + \tilde{e}_{0m,t} + O_p \left( \frac{1}{C_{2NT}} \right) = \sigma_{m,t}^2 + O_p \left( \frac{1}{\sqrt{N}} \right), \]

where \( C_{2NT} = \min\{N, \sqrt{T}\} \) and \( \tilde{e}_{0m,t} = \sum_{i=N_{m-1}+1}^{N_m} (e_{i,t}^2 - \sigma_{m,t}^2)/(N_m - N_{m-1}) \). This, together with Taylor expansion of the type \( \log(1+x) = x + O(x^2) \), gives

\[ \log \left( \frac{\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2} \right) = \log \left( 1 + \frac{\hat{\sigma}_{m,t}^2 - \sigma_{m,t}^2}{\sigma_{m,t}^2} \right) = \frac{\hat{\sigma}_{m,t}^2 - \sigma_{m,t}^2}{\sigma_{m,t}^2} + O_p((\hat{\sigma}_{m,t}^2 - \sigma_{m,t}^2)^2) \]

\[ = \frac{\hat{\sigma}_{m,t}^2}{\sigma_{m,t}^2} - 1 + O_p(e_{0m,t}^2), \]

where \( \hat{\sigma}_{m,t}^2/\sigma_{m,t}^2 - 1 = \tilde{e}_{0m,t}/\sigma_{m,t}^2 = O_p(1/\sqrt{N}) \) and \( O_p(\tilde{e}_{0m,t}^2) = O_p(1/N) \). Hence, since \( \hat{\sigma}_{m,t}^2/\sigma_{m,t}^2 - 1 \) is dominated by \( O_p(\tilde{e}_{0m,t}^2) \), we can show that

\[ QML(n) - QML_0(N_1) \]

\[ \sim \sum_{t=p+1}^{T} \left[ \frac{\hat{\sigma}_{1,t}^2(n)}{\sigma_{1,t}^2} + (N-n) \frac{\hat{\sigma}_{2,t}^2(n)}{\sigma_{2,t}^2} - N + (n-N_1) \log \left( \frac{\sigma_{1,t}^2}{\sigma_{2,t}^2} \right) \right]. \quad (A50) \]

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 \( n \),

\[ \hat{\sigma}_{1,t}^2(n) = \frac{1}{n} \sum_{i=1}^{n} \epsilon_{i,t}^2 + O_p \left( \frac{1}{C_{2NT}} \right) = \sigma_{1,t}^2 \frac{1}{n} \sum_{i=1}^{n} \epsilon_{i,t}^2 + O_p \left( \frac{1}{C_{2NT}} \right), \]

and

\[ \hat{\sigma}_{2,t}^2(n) = \frac{1}{(N-n)} \sum_{i=n+1}^{N} \epsilon_{i,t}^2 + O_p \left( \frac{1}{C_{2NT}} \right) \]

\[ = \frac{1}{(N-n)} \left( \sigma_{1,t}^2 \sum_{i=n+1}^{N} \epsilon_{i,t}^2 + \sigma_{2,t}^2 \sum_{i=N_{t+1}}^{N} \epsilon_{i,t}^2 \right) + O_p \left( \frac{1}{C_{2NT}} \right), \]

49
from which we obtain

\[ QML(n) - QML_0(N_1) \]

\[ \sim \sum_{t=p_t+2}^{T} \left[ n \epsilon_{i,t}^2 + \frac{\sigma_i^2}{\sigma_{i,t}^2} \sum_{i=n+1}^{N_1} \epsilon_{i,t}^2 + \sum_{i=N_1+1}^{N} \epsilon_{i,t}^2 - N + (n - N_1) \log \left( \frac{\sigma_i^2}{\sigma_{i,t}^2} \right) \right] \]

\[ = \sum_{t=p_t+2}^{T} \left[ \sum_{i=1}^{N} \epsilon_{i,t}^2 + x_i \sum_{i=n+1}^{N_1} \epsilon_{i,t}^2 - N + (n - N_1) \log(1 + x_i) \right] \]

\[ = \sum_{t=p_t+2}^{T} \left[ \sum_{i=1}^{N} (\epsilon_{i,t}^2 - 1) + \sum_{i=n+1}^{N_1} [x_i \epsilon_{i,t}^2 - \log(1 + x_i)] \right], \]

where \( x_i = \sigma_i^2 / \sigma_{i,t}^2 - 1 \). Hence, since

\[ \sum_{i=n+1}^{N_1} [x_i \epsilon_{i,t}^2 - \log(1 + x_i)] = \sum_{i=n+1}^{N_1} [x_i \epsilon_{i,t}^2 - \log(1 + x_i)] \]

\[ = (N_1 - n)[x_i - \log(1 + x_i)] + \sum_{i=n+1}^{N_1} x_i (\epsilon_{i,t}^2 - 1), \]

we can show that

\[ \frac{1}{\sqrt{NT}} [QML(n) - QML_0(N_1)] \]

\[ \sim \frac{1}{\sqrt{NT}} \sum_{t=p_t+2}^{T} \sum_{i=1}^{N} (\epsilon_{i,t}^2 - 1) + \sqrt{NT} \frac{(N_1 - n)}{NT} \sum_{t=p_t+2}^{T} [x_i - \log(1 + x_i)] \]

\[ + \frac{1}{\sqrt{NT}} \sum_{i=n+1}^{N_1} \sum_{t=p_t+2}^{T} x_i (\epsilon_{i,t}^2 - 1). \] (A51)

From here it is easy to see that if \( n = N_1 \), then

\[ \frac{1}{\sqrt{NT}} [QML(n) - QML_0(N_1)] \sim \frac{1}{\sqrt{NT}} \sum_{t=p_t+2}^{T} \sum_{i=1}^{N} (\epsilon_{i,t}^2 - 1) = O_p(1), \] (A52)

whereas if \( n < N_1 \), then

\[ \frac{1}{\sqrt{NT}} [QML(n) - QML_0(N_1)] \sim \sqrt{NT} \frac{(N_1 - n)}{NT} \sum_{t=p_t+2}^{T} [x_i - \log(1 + x_i)] + O_p(1) \]

\[ = O_p(\sqrt{NT}). \] (A53)

Note that \( QML_0(N_1) \) is independent of \( n \). Hence, minimizing \( QML(n) \) is equivalent to minimizing \( [QML(n) - QML_0(N_1)] \). Because of this, and the fact that the order of magnitude of \( [QML(n) - QML_0(N_1)] \) 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. ■
Table 1: Size at the 5% level.

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>Case</th>
<th>$\tau$</th>
<th>$\bar{\sigma}_{IV}$</th>
<th>$\delta_{NT}$</th>
<th>$\tau_i$</th>
<th>$\tau_{IV,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>7.0</td>
<td>0.8</td>
<td>3.0</td>
<td>5.3</td>
<td>1.3</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>1</td>
<td>5.6</td>
<td>0.1</td>
<td>3.9</td>
<td>5.1</td>
<td>0.4</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2</td>
<td>7.7</td>
<td>1.4</td>
<td>2.7</td>
<td>5.6</td>
<td>2.2</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>2</td>
<td>5.6</td>
<td>0.3</td>
<td>3.7</td>
<td>5.2</td>
<td>0.8</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3</td>
<td>8.0</td>
<td>1.5</td>
<td>3.0</td>
<td>5.6</td>
<td>2.3</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>3</td>
<td>5.6</td>
<td>0.3</td>
<td>3.8</td>
<td>5.2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

$\phi_1 = 0$

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>Case</th>
<th>$\tau$</th>
<th>$\bar{\sigma}_{IV}$</th>
<th>$\delta_{NT}$</th>
<th>$\tau_i$</th>
<th>$\tau_{IV,i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>8.4</td>
<td>0.0</td>
<td>3.0</td>
<td>5.0</td>
<td>0.3</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>1</td>
<td>5.9</td>
<td>0.0</td>
<td>3.6</td>
<td>5.0</td>
<td>0.1</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2</td>
<td>9.3</td>
<td>0.2</td>
<td>3.1</td>
<td>5.5</td>
<td>0.6</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>2</td>
<td>6.3</td>
<td>0.0</td>
<td>4.0</td>
<td>5.2</td>
<td>0.2</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3</td>
<td>9.5</td>
<td>0.2</td>
<td>3.1</td>
<td>5.5</td>
<td>0.7</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>3</td>
<td>6.4</td>
<td>0.0</td>
<td>4.1</td>
<td>5.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

$\phi_1 = 0.5$

Notes: Variance cases 1–3 refer to homogeneity, a discrete break and a smooth transition break, respectively. $\phi_1$ refers to the first-order autoregressive serial correlation coefficient in the errors.
Table 2: Power at the 5% level

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>Case</th>
<th>$\tau$</th>
<th>$\bar{\tau}_{IV}$</th>
<th>$\bar{S}_{NT}$</th>
<th>$\tau_i$</th>
<th>$\tau_{IV,i}$</th>
<th>$\kappa = 0$</th>
<th>$\kappa = 1/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>31.8</td>
<td>3.6</td>
<td>8.1</td>
<td>6.2</td>
<td>2.0</td>
<td>11.1</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>1</td>
<td>64.6</td>
<td>1.1</td>
<td>14.6</td>
<td>6.4</td>
<td>0.5</td>
<td>10.3</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2</td>
<td>45.4</td>
<td>5.9</td>
<td>7.2</td>
<td>6.4</td>
<td>3.4</td>
<td>13.8</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>2</td>
<td>88.0</td>
<td>2.3</td>
<td>13.9</td>
<td>6.6</td>
<td>1.1</td>
<td>14.7</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3</td>
<td>45.8</td>
<td>6.2</td>
<td>7.4</td>
<td>6.5</td>
<td>3.4</td>
<td>14.1</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>3</td>
<td>88.1</td>
<td>2.3</td>
<td>14.4</td>
<td>6.6</td>
<td>1.1</td>
<td>14.7</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: $\kappa$, $a$, and $b$ are such that $\rho_i = \exp(c_i/N^\kappa T)$, where $c_i \sim U(a, b)$. See Table 1 for an explanation of the rest.
Table 3: Power at the 5% level

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>Case</th>
<th>$\kappa = 0$</th>
<th>$\kappa = 1/2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\tau$</td>
<td>$\hat{\tau}_{IV}$</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>97.1</td>
<td>38.5</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>1</td>
<td>100.0</td>
<td>39.1</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2</td>
<td>92.6</td>
<td>34.8</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>2</td>
<td>100.0</td>
<td>32.4</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3</td>
<td>93.6</td>
<td>35.2</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>3</td>
<td>100.0</td>
<td>32.5</td>
</tr>
</tbody>
</table>

Notes: See Tables 1 and 2 for an explanation.