POOLED PANEL UNIT ROOT TESTS AND THE EFFECT OF PAST INITIALIZATION∗

Joakim Westerlund†
Deakin University
Australia

March 15, 2013

Abstract

This paper analyzes the role of initialization when testing for a unit root in panel data, an issue that has received surprisingly little attention in the literature. In fact, most studies assume that the initial value is either zero or bounded. As a response to this, the current paper considers a model in which the initialization is in the past, which is shown to have several distinctive features that makes it attractive, even in comparison to the common time series practice of making the initial value a draw from its unconditional distribution under the stationary alternative. The results have implications not only for theory, but also for applied work. In particular, and in contrast to the time series case, in panels the effect of the initialization need not be negative but can actually lead to improved test performance.

JEL Classification: C22; C23.

Keywords: Panel unit root test; Initial value; Local asymptotic power.

1 Motivation

Consider the panel data variable $y_{i,t}$, observable for $t = 1, ..., T$ time series and $i = 1, ..., N$ cross-sectional units. While the literature concerned with the analysis of unit roots in such variables is huge and covers more than 20 years, the role of the initial value, $y_{i,0}$, is hardly

∗The author would like to thank the editor and two anonymous referees for many valuable comments and suggestions. The author would also like to thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant numbers P2005–0117:1 and W2006–0068:1.
†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.
ever discussed. This is especially noteworthy given well-known initial value effect in the
time series case. In fact, the size of the initial value strongly influences the performance of
time series unit root tests, up to the point of reversing the ranking of different tests (see, for
example, Stock, 1994). In practice this means that different conclusions can be reached with
samples of the same data that differ only in the date at which the sample begin (see Elliott
and Müller, 2006, for an illustration). It also means that the decision of which test to choose
is not trivial, as the researcher is forced to take a stand on the size of the initial value, which
is neither known nor is it amenable to estimation. The decision therefore comes down to
ones prior beliefs, and in most situations it is hard to rule out small or large initial values
(see Elliott and Müller, 2006).

Most panel studies assume that the initial value is either a fixed constant, typically set
to zero, or drawn from a bounded distribution. The reason for keeping with this rather
unrealistic initial value assumption is that it is convenient; if $y_{i,0}$ is bounded, as far as the
asymptotic theory is concerned, it can be ignored. Relaxing the bounded initial value as-
sumption also creates a need to be explicit about the allowable initializations, and it is not
obvious how to do this.

In the time series case, as an alternative to the bounded initial value assumption, it is
common to assume that the initial value is drawn from its unconditional distribution under
the stationary alternative (see, for example, Elliott, 1999; Elliott and Müller, 2006; Müller
and Elliott, 2003), which is quite plausible, as the beginning of the sample is unlikely to
coincide with the beginning of the process under study. In fact, one way to think about
this assumption is as if the process has been running for some time prior to the start of the
sample. In the local-to-unity setting with an autoregressive root that shrinks towards the null
at the rate $1/T$, this means the initial value is $O_p(\sqrt{T})$, which is very appealing in the sense
that the magnitude of the initial value is the same as that of the observed data under the
unit root null. In other words, if the observed time series wanders according to a stochastic
trend, it seems fair to assume that the initialization itself may be regarded as the outcome of
a similarly random wandering process.

Unfortunately, the appeal of the unconditional distribution assumption does not extend
to the panel context. The reason is that if $y_{i,0}$ is drawn from its unconditional distribution
under the usual panel local alternative in which the rate of shrinking is given by $1/\sqrt{NT}$,
then $y_{i,0} = O_p(N^{1/4}/\sqrt{T})$, suggesting that the magnitude of the initial condition for each unit
should depend on the total number of units, $N$, which is of course not very realistic. Harris et al. (2010) consider the local power of the Im et al. (2003) panel unit root test under this very assumption. As a rationale they state (on page 313): “we do not consider our specification to necessarily be empirically realistic (that is, the implication that the initial condition depends on the cross-sectional dimension has questionable plausibility), but is instead chosen simply for its suitability in the asymptotic analysis.” Moon et al. (2007, page 436) discuss the issue of initialization in the panels. Their conclusion is:

This example [the fact that in panels the unconditional distribution assumption implies that $y_{i,0}$ should grow with $N$] makes it clear that mechanical extensions of time series formulations that are commonly used for initial conditions can lead to quite unrealistic and unjustifiable features in a panel context. It is therefore necessary to consider initializations that are sensible for panel models, while at the same time having realistic time series properties.

They end (again on page 436) by stating that: “it is an important matter for future research to extend the theory and relax this condition”. This paper can be seen as a step in this direction. Specifically, rather than assuming that $y_{i,0}$ drawn from its unconditional distribution under the local alternative, we assume that the initialization takes place somewhere in the past, which is the same idea as in Phillips et al. (2001), Phillips and Magdalinos (2009), and Andrews and Guggenberger (2008), who study the effect of initialization in the pure time series case. But while the main attraction of past initialization in time series is that the reflects the idea that the process has been running for some time prior to the start of the sample, in the panel setting it has the additional attraction of breaking the link between the magnitude of the initial condition and the rate of shrinking of the local alternative, thereby making it a natural alternative to the unconditional distribution assumption.

But it is not only the motivation that differs. In fact, the asymptotic results obtained in the current panel setting are materially different from those obtained in the pure time series case, which we illustrate using as an example the usual pooled ordinary least squares (OLS) $t$-statistic for a unit root in a simple but transparent model with time series and cross-section correlation free errors. Our main findings can be summarized as follows. First, in contrast to the time series case where the asymptotic null distribution depends critically on the extent of the initialization (as measured by how far in the past the initial value is allowed to reach), the asymptotic null distribution of the pooled unit root $t$-statistic is always standard normal,
even if fixed effects are not allowed for. Thus, if size accuracy is the only concern, under
the assumptions of the paper, one does not need to bother about the initial value. Second,
without fixed effects power is generally increasing in the extent of the initialization, which is
again different from the time series case where power is generally decreasing in the extent of
the initialization. Third, unless in the infinite past, if fixed effects are included the $t$-statistic
considered here is asymptotically invariant with respect to the initialization. This is in con-
trast to what happens in time series where power is again decreasing in the initialization,
even in models with fitted intercept and trend terms.\footnote{While under the null hypothesis time series tests (with at least an intercept) are invariant with respect to the initialization, this is not the case under the alternative (see Elliott and Müller, 2006).}

The results reported here differ not only from what might be expected based on the re-
lated time series literature, but stand out also against other panel data studies. The most
noticeable difference is, of course, that, unlike most previous work, in this study the initial-
ization is given a formal treatment. In fact, as far as we are aware the only other panel study
that does not assume that $y_{i,0}$ is either zero or bounded is that of Harris et al. (2010). How-
ever, they assume that $y_{i,0}$ is drawn from its unconditional distribution, which we have seen
leads to the rather unrealistic prediction that the effect of the initialization should grow with
$N$. Second, while most existing power functions are mere first-order approximations, our
analysis is based on an expansion of the test statistic that keep terms that are of higher order
in the magnitude, and is therefore expected to lead to better predictions in small samples, a
result confirmed by our simulations.

2 Model discussion

To be able to capture the distant past initialization, we assume that

\begin{align}
y_{i,t} &= \beta_i + u_{i,t}, \\
u_{i,t} &= \rho_i u_{i,t-1} + \epsilon_{i,t},
\end{align}

where $t = -T_0 + 2, \ldots, T$, and $\epsilon_{i,t}$ is independently and identically distributed (iid) across
$(i,t)$ with $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = \sigma^2 > 0$ and $E(\epsilon_{i,t}^4) < \infty$. The fixed effect $\beta_i$ need not be
present, which means that we will be considering two deterministic models; (i) $\beta_i = 0$, and
(ii) $\beta_i$ unrestricted. The process is initiated at $u_{i,-T_0+1} = 0$ and the relative extent of the
initialization is governed by $\tau_T = T_0/T$, where we assume for simplicity that $T_0 = T_0(T) =$
$T^\kappa$, where $\kappa \geq 0$, which in turn implies $\tau_T = T^{\kappa-1}$. Hence, by determining $\kappa$ we can control the value of $\tau_T$. There are three cases;

(i) $\kappa \in [0, 1)$,

(ii) $\kappa = 1$, and

(iii) $\kappa > 1$,

henceforth referred to as “recent past”, “distant past” and “infinite past” initialization, respectively (Phillips and Magdalinos, 2009). In order to appreciate fully the distinction between these cases, it is convenient to rewrite (2) as

$$u_{i,t} = \sum_{s=0}^{T_0-2} \rho^{l+s}_i \epsilon_{i,-s} + \sum_{s=1}^{l} \rho^{l-s}_i \epsilon_{i,s} = \rho^l_i u_{i,0} + \sum_{s=1}^{l} \rho^{l-s}_i \epsilon_{i,s},$$

where $u_{i,0} = \sum_{s=0}^{T_0-2} \rho^s_i \epsilon_{i,-s}$ is the initial value, which is of order $O_p(\sqrt{T_0}) = O_p(T^{\kappa/2})$ (see Appendix). An alternative way of thinking about $\kappa$ is therefore that it controls the relative order of the initial value effect, $\rho^l_i u_{i,0}$, when compared to the (near) random walk process $\sum_{s=1}^{l} \rho^{l-s}_i \epsilon_{i,s}$, whose order is given by $O_p(\sqrt{T})$. If $\kappa = 1$, then the initial value is of the same order as that of the random walk, whereas if $\kappa > 1$, then the initial value will tend to dominate. If $\kappa \in [0, 1)$, then $O_p(T^{\kappa/2}) = o_p(\sqrt{T})$, and therefore the initial value will be dominated by the random walk.

Similarly to the assumption placed on $T_0$ and $T$, in order to control the relative expansion rate of $N$ and $T$, we assume that $T = T(N) = N^\theta$, where $\theta > 0$. This means that $\tau_T$ can be further rewritten as $\tau_T = \tau_N = N^{\theta(k-1)}$.

While simple, as we will show later, the above model is able to deliver significant insight. It also allows us to put more focus on the initialization, which is going to depend in an intricate way on the persistency of $u_{i,t}$, as measured by $\rho_i$. In order to capture this, the following local-to-unity model will be used:

$$\rho_i = \exp \left( \frac{\alpha_N c_i}{T} \right),$$

where $\alpha_N = 1/N^\eta$, $\eta \geq 0$ and the drift parameter $c_i$ is assumed to be iid and independent of $\epsilon_{i,t}$. All moments of $c_i$ exist, and in what follows it will be convenient to denote these as $\mu_m = E(\epsilon_i^m)$ for $m \geq 0$ with $\mu_0 = 1$. $\theta$, $\kappa$ and $\eta$ are related via $\theta(\kappa - 1) - \eta \leq 0$, which is mainly a technical condition (see remark 2 below). However, except for this and the requirement...
that $\kappa, \eta \geq 0$ and $\theta > 0$, nothing is assumed a priori regarding the values taken by these parameters, which are instead considered a part of the analysis. This stands in sharp contrast to existing studies where $\eta = 1/2$ is typically assumed from the outset (see, for example, Breitung, 2000).

The null hypothesis that we will consider is that of a unit root, which can be written as $H_0 : c_1 = \ldots = c_N = 0$, or equivalently, $H_0 : \mu_2 = 0$. The formulation of the alternative hypothesis depends on what one is willing to assume regarding the sign and homogeneity of $c_i$. Here we make no assumptions and therefore the alternative is formulated simply as $H_1 : c_i \neq 0$ for some $i$, or $H_1 : \mu_2 > 0$. Thus, while some units may be “locally stationary” ($c_i < 0$), others might be “locally explosive” ($c_i > 0$). There is also nothing to prevent some (but not all) of the units from being unit root non-stationary ($c_i = 0$).

**Remarks.**

1. The above model is general enough to nest most local alternatives and initializations considered so far in the literature. Consider the local alternative in (3). By considering $\eta = \{1/4, 1/2\}$, we cover almost the entire panel unit root literature, and by adding $\eta = 0$, we cover also the time series literature. As for the initialization, by setting $\kappa = 1$, the order of the initial value is the same as under the unconditional distribution assumption in the time series case. Our model also covers the panel version of this assumption. To illustrate this, suppose for simplicity that $\theta = 1$, such that $T = N$, in which case the rate of expansion of the initial value under the panel unconditional distribution assumption is given by $N^{1/4} \sqrt{T} = N^{3/4}$. The same rate can be obtained in the current model by simply solving $T^{\kappa/2} = N^{\kappa/2} = N^{3/4}$ for $\kappa$, giving $\kappa = 3/2$. Finally, if $\kappa = 0$, then $T_0 = 1$, and so we are back in the fixed initial value setup with $u_{i,-T_0+1} = u_{i,0} = 0$.

2. Many of the above assumptions are not necessary and can be relaxed at the expense of added technical complexity, which is unnecessary in the present case, because the test statistic that we will consider has already been extended to accommodate more general data generating processes. The assumption that $\epsilon_{it}$ is homoskedastic and independent through time can, for example, be relaxed in a relatively straightforward manner (see, for example, Westerlund and Blomquist, 2012). Relaxing the assumed independence across the cross-section is more involved, although factor model approaches have been
shown to work (see, for example, Bai and Ng, 2010; Moon and Perron, 2004). Similarly, the requirement that \( u_{i,t} \) is initiated at zero can be relaxed by assuming that \( u_{i,-T_0+1} = O_p(1) \).\(^2\) The assumption that \( c_i \) has all its moments is also not necessary, but is made here in order to keep with the rest of the local asymptotic literature (see, for example, Moon et al., 2007). That \( c_i \) need not have finite moments is particularly obvious under the null, in which case \( c_i \) and all its moments are zero. The condition that \( \theta(\kappa - 1) - \eta \leq 0 \) is only binding when \( \kappa > 1 \). In order to appreciate the need for this requirement, note first that the slowest rate of shrinking of the local alternative possible is obtained by setting \( \eta = 0 \), such that \( \alpha_N = 1 \) and therefore \( \rho_i = \exp(\alpha_N c_i / T) = \exp(c_i / T) \), which is nothing but the conventional local-to-unity specification in the time series case (see Andrews and Guggenberger, 2008, for a discussion). The rate of shrinking must therefore be at least \( 1 / T \), which is satisfied if \( \kappa \in [0,1] \) (see Appendix). However, if \( \kappa > 1 \), then the appropriate “drift” of the initial value is no longer given by \( \alpha_N c_i \), but rather by \( \alpha_N \tau_N c_i \). Hence, in order to ensure that the rate of shrinking is at least \( 1 / T \), we need \( \alpha_N \tau_N = N^{\theta(\kappa - 1) - \eta} = O(1) \), which is satisfied if \( \theta(\kappa - 1) - \eta \leq 0 \).

3. Suppose for simplicity that \( \kappa \in [0,1] \). In this case, the condition that \( T_0 = T^\kappa \) implies \( \tau_N \to \{0,1\} \), which is less “flexible” than the setup of Phillips and Magdalinos (2009), in which \( T_0 / T \to \tau \in [0,1] \). However, since the conclusions are qualitatively the same, and since assuming \( T_0 = T^\kappa \) greatly simplifies both transparency and notation, in the present paper we opt for the less flexible specification. Similarly, while \( T \sim N^\theta \) is relatively more “flexible”, \( T = N^\theta \) is more transparent. In the terminology of Phillips and Moon (1999), we assume a “diagonal path” relationship between \( T_0 \), \( T \) and \( N \). Under this scheme, in order to pass all three indices (\( T_0 \), \( T \) and \( N \)) to infinity, it is enough to let \( N \to \infty \).

\(^2\)In many testing situations there are no reasons to expect the initial observation to be “unusual”, where unusual starting observations imply an unbounded initial condition. Therefore, in situations like this the assumption that \( u_{i,-T_0+1} = O_p(1) \) is not very restrictive.
3 Asymptotic results

3.1 No fixed effects

A natural candidate for a unit root test statistic when \( \beta_i = 0 \) is given by the following pooled OLS t-statistic:

\[
t = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta y_{it} y_{i,t-1} = \frac{A_N}{\hat{\sigma} \sqrt{B_N}},
\]

with obvious definitions of \( A_N \) and \( B_N \) (which are only subscripted by \( N \), because \( T \) is a function of \( N \)), and \( \hat{\sigma}^2 \) being any consistent estimator of \( \sigma^2 \) satisfying \( \sqrt{N}(\hat{\sigma}^2 - \sigma^2) = o_p(1) \).\(^3\)

The reason for considering this particular test statistic is in part because of its popularity in the panel unit root literature (see, for example, Levin et al., 2002; Moon and Perron, 2008; Moon et al., 2007), in part because it has been used as a basis for numerous extensions, including tests for cointegration (see Westerlund and Breitung, 2012).

By using Taylor expansion of the type \( \rho_i = \exp(\alpha N c_i / T) = 1 + \alpha N c_i / T + O_p(\alpha^2 N / T^2) \) and then (2), it is clear that \( A_N \) can be rewritten as

\[
A_N = \sqrt{N}A_{1N} + A_{2N} + O_p \left( \frac{\sqrt{N}A_N^2}{T^2} \right),
\]

where

\[
A_{1N} = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} c_i y_{i,t-1}^2,
\]

\[
A_{2N} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t} \epsilon_{i,t}.
\]

Lemma 1 provides the limits for \( A_{1N}, A_{2N} \) and \( B_N \).

**Lemma 1.** Under \( \beta_i = 0 \) and the conditions laid out in Section 2, as \( N \to \infty \),

(a) \( A_{1N} = \sigma^2(\lambda_{1N} + \tau_N \gamma_{1N}) + o_p(1) \),

(b) \( A_{2N} \to_d \sigma^2 \left( \lim_{N \to \infty} (\lambda_{0N} + \tau_N \gamma_{0N}) \right)^{1/2} N(0,1) \),

(c) \( B_N = \sigma^2(\lambda_{0N} + \tau_N \gamma_{0N}) + o_p(1) \),

\(^3\)See Section 4 for an example of how \( \hat{\sigma}^2 \) might be constructed.
where \( \to_d \) signifies convergence in distribution and

\[
\begin{align*}
\lambda_{pN} &= \sum_{j=0}^{\infty} \phi_{jN} \mu_{j+p}, \\
\gamma_{pN} &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (j+2)(k+2) \tau_{N} \phi_{jN} \phi_{kN} \mu_{k+j+p}, \\
\phi_{jN} &= \left( \frac{2\alpha_N}{j+2} \right)!.
\end{align*}
\]

From Lemma 1 we can deduce easily the limiting distribution of \( t \). Indeed, by using (4) and the consistency of \( \hat{\sigma}^2 \), provided that \( \eta + \theta \geq 1/4 \), such that \( \sqrt{N\alpha_N^2/T^2} = N^{1/2-2(\eta+\theta)} = o(1) \) (only required under \( H_1 \)), we have

\[
t = \sqrt{N\alpha_N} \frac{A_{1N}}{\sigma\sqrt{B_N}} + \frac{A_{2N}}{\sigma\sqrt{B_N}} + o_p(1), \tag{5}
\]

where

\[
\begin{align*}
\frac{A_{1N}}{\sigma\sqrt{B_N}} &= \frac{\lambda_{1N} + \tau_N \gamma_{1N}}{\sqrt{\lambda_{0N} + \tau_N \gamma_{0N}}} + o_p(1), \tag{6} \\
\frac{A_{2N}}{\sigma\sqrt{B_N}} &\to_d N(0,1) \tag{7}
\end{align*}
\]

as \( N \to \infty \). Thus, if the null hypothesis is true so that \( c_i \) and all its moments are zero, then \( \lambda_{1N} = \gamma_{1N} = 0 \), and therefore

\[
t = \frac{A_{2N}}{\sigma\sqrt{B_N}} + o_p(1) \to_d N(0,1).
\]

Because this result holds independently of \( \tau_N \), any dependence on the initial value must come from power, as captured by the first term on the right-hand side of (5), which also determines the mean of the test statistic. Making use of (6), this term can be written as

\[
\sqrt{N\alpha_N} \frac{A_{1N}}{\sigma\sqrt{B_N}} = \sqrt{N\alpha_N} \frac{\lambda_{1N} + \tau_N \gamma_{1N}}{\sqrt{\lambda_{0N} + \tau_N \gamma_{0N}}} + o_p(\sqrt{N\alpha_N}),
\]

where \( \eta \geq 1/2 \) ensures that the remainder is \( o_p(\sqrt{N\alpha_N}) = o_p(N^{1/2-\eta}) = o_p(1) \). It follows that if \( \kappa \in [0,1) \), since \( \theta > 0 \), \( \tau_N = N^{\theta(k-1)} \to 0 \), and therefore the initial value effect, captured by \( \tau_N \gamma_{1N} \) and \( \tau_N \gamma_{0N} \), disappears. If, on the other hand, \( \kappa = 1 \) such that \( \tau_N = 1 \), then \( \tau_N \gamma_{1N} \) and \( \tau_N \gamma_{0N} \) are of the same order as \( \lambda_{0N} \) and \( \lambda_{1N} \) (which do not depend on the

\[\text{4} \] The order of the remainder terms in Lemma 1 (a) and (c) is not the sharpest possible. Hence, from this point of view, the assumption that \( \eta \geq 1/2 \) is probably stronger than necessary. However, since the same assumption is needed to ensure a meaningful analysis of local power (see remark 7), there seem to be little or no point in trying to relax it here.
initialization), and therefore power will depend on the initialization. Finally, if $\kappa > 1$, then $\tau_N \to \infty$, and so

$$\sqrt{N}a_N \frac{\lambda_{1N} + \tau_N \gamma_{1N}}{\sqrt{\lambda_{0N} + \tau_N \gamma_{0N}}} = \sqrt{N} \tau_N a_N \left[ \frac{\gamma_{1N}}{\sqrt{\gamma_{0N}}} + O \left( \frac{1}{\tau_N} \right) \right],$$

suggesting that now the initial value has a dominating effect (with $\lambda_{0N}$ and $\lambda_{1N}$ being absorbed by the $O(1/\tau_N)$ reminder term).

Remarks.

4. Lemma 1 generalizes the previous work on the local power of panel unit root tests in two directions. Firstly, it shows how $t$ is affected by past initialization, an issue that has not been considered before. Secondly, while most research assume that $a_N = o(1)$ and only report results for the resulting first-order approximate power function (see, for example, Moon et al., 2007), which only depends on $\mu_1$, Lemma 1 accounts for all the moments of $c_i$ and is therefore expected to produce more accurate predictions, a result that is verified using Monte Carlo simulation in Section 4.\(^5\)

5. Previous works by Breitung (2000, Theorem 4) and Moon et al. (2007, Section 3.3.1) have shown that under the null hypothesis and in the special case when $\kappa = 0$, $t \to_d N(0,1)$. Lemma 1 shows that the same result applies even when the initialization is in the past. The fact that the same result applies regardless of what is being assumed regarding $\kappa$ stands in sharp contrast to the pure time series case, in which the validity of the conventional unit root theory requires $T_0/T \to 0$ (see Phillips and Magdalinos, 2009). In other words, even when applied directly to the raw data (without demeaning) the asymptotic size of the test is independent of the initialization. Panel data tests are therefore more robust in this regard. This a great advantage, because it means that the appropriate critical values to use are the same regardless of the value of $\kappa$, which in practice is of course unknown.

6. While the asymptotic size of the test does not depend on the initialization, power does. The fact that the results differ depending on whether $\kappa \in [0,1)$, $\kappa = 1$ or $\kappa > 1$ is in agreement with the time series results of Andrews and Guggenberger (2008, Proposition 1). The way in which this difference materializes itself is, however, very different.

\(^5\)In fact, when it comes to the accuracy of the approximation, the only study that comes close is that of Westerlund and Larsson (2012), and then only the first four moments of $c_i$ are accounted for.
In Andrews and Guggenberger (2008), while $\kappa \in [0, 1]$ leads to a conventional unit root type distribution, $\kappa > 1$ leads to a Cauchy distribution. In our case, the distribution is always normal but with a differing mean. If $\kappa \in [0, 1)$, the mean is given by $\sqrt{N\alpha N\lambda_{1N}}/\sqrt{\lambda_{0N}}$, if $\kappa = 1$, it is given by $\sqrt{N\alpha N}(\lambda_{1N} + \tau_N\gamma_{1N})/\sqrt{\lambda_{0N} + \tau_N\gamma_{0N}}$, and if $\kappa > 1$, it is given by $\sqrt{N\alpha N\gamma_{1N}}/\sqrt{\gamma_{0N}}$ (as shown in the discussion following Lemma 1). Note that $\gamma_{1N} = \mu_1 + O(\alpha N)$, where in case of a locally stationary alternative, $\mu_1 < 0$. This means that if $\kappa \geq 1$ a larger $\tau_N$ is going to make the numerator of the mean of the test statistic more negative, which is suggestive of higher power. This is in contrast to the time series case in which power is typically decreasing in the size of the initial condition (see, for example, Elliott and Müller, 2006).\(^6\) Hence, even under the alternative the use of panel data leads to a measure of robustness that is not there in time series. If explosiveness is permitted, such that $\mu_1 > 0$ cannot be ruled out, since this is going to push the mean of the test statistic in the other (positive) direction, we recommend making the test double-sided.

7. As pointed out in remark 6 above, power depends on whether $\kappa \in [0, 1), \kappa = 1$ or $\kappa > 1$. However, the most important difference is between the cases of $\kappa \in [0, 1]$ and $\kappa = 1$. Consider first the case when $\kappa \in [0, 1]$. If $\eta > 1/2$, such that $\sqrt{N\alpha N} = N^{1/2-\eta} \rightarrow 0$, then power is negligible, whereas if $\eta < 1/2$, such that $\sqrt{N\alpha N} \rightarrow \infty$, then $t$ diverges and therefore power goes to one as $N \rightarrow \infty$. Only in the intermediate case when $\eta = 1/2$, such that $\sqrt{N\alpha N} = 1$, is power non-negligible and non-increasing. The corresponding condition for non-negligible and non-increasing power when $\kappa > 1$ is given by $\sqrt{N\alpha N} = N^{\theta(\kappa-1)/2-\eta+1/2} = 1$, which is satisfied if $\theta(\kappa-1)/2-\eta+1/2 = 0$. Hence, since $\theta > 0$, for this to hold, we need $\eta = (\theta(\kappa-1) + 1)/2 > 1/2$, which means that $\eta = 1/2$ (the standard value in the literature) is effectively ruled out as being too large for non-increasing power. That is, setting $\eta = 1/2$ causes power to go to one with the sample size. In other words, if we compare the two cases of $\kappa \in [0, 1]$ and $\kappa > 1$, the latter actually implies an increase in the $1/N\eta T$-neighborhoods around unity for which power is non-negligible, meaning that in this case we can be even closer to the null than before and still have power.

8. The fact that when $\kappa > 1$ and $\eta = 1/2$ power goes to one with the sample size (see

\(^6\)See Harris et al. (2010) for a similar result in the panel data context.
remark 7) is a reflection of the fact that the rate of convergence of the pooled OLS slope estimator in a regression of $\Delta y_{i,t}$ onto $y_{i,t-1}$ is now given by $\sqrt{N}t_N$ (see Appendix), which is faster than the usual $\sqrt{NT}$ rate. This is consistent with the results reported by Andrews and Guggenberger (2008, Theorem 1), and Phillips and Magdalinos (2009, Theorem 2) for the time series case.

9. $\eta$ and $\kappa$ determine not only the extent of power (see remark 7), but also how power is affected by the moments of $c_i$. There are two cases, depending on whether $\kappa \in [0,1]$ or $\kappa > 1$. If $\kappa \in [0,1]$, such that $t_N \to \{0,1\}$, since $t_N\gamma_{1N}$ and $t_N\gamma_{0N}$ are either asymptotically zero or independent of $t_N$, the effect of the moments of $c_i$ is determined by $\eta$ alone. On the one hand, if $\eta = 0$, such that $\sigma_N = 1$, then $\phi_{1N} = (2\alpha_N)^j/(j+2)! = O(1)$, and therefore all the moments of $c_i$ affect power (via $\lambda_{pN}$ and $\gamma_{pN}$). On the other hand, if $\eta > 0$, such that $\sigma_N = o(1)$, then $\phi_{1N} = O(\alpha_N^j) = o(1)$ and therefore $\lambda_{1N} = \mu_1/2 + o(1)$ and $\gamma_{1N} = \mu_1 + o(1)$, suggesting that now $\mu_1$ is going to have a dominating effect. For instance, if $\eta = 1/2$ and $\kappa = 0$, then $t_1 \to_d \mu_1/\sqrt{2} + N(0,1)$, which is the same result as in, for example, Breitung (2000, Theorem 2) and Moon et al. (2007, Section 3.3.1). This previous result can therefore be seen as a special case of the more general theory developed here. If $\kappa > 1$, then the effect of the moments of $c_i$ is determined by both $\eta$ and $\kappa$.

3.2 Fixed effects

Let $\bar{y}_i = \sum_{t=1}^{T} y_{i,t} / T$ and let $A_{1N}^*, A_{2N}^*$ and $B_N^*$ be defined as $A_{1N}, A_{2N}$ and $B_N$, respectively, but with $y_{i,t}^* = (y_{i,t} - \bar{y}_i)$ in place of $y_{i,t}$. Lemma 2 provides the relevant asymptotic results for the case when $\beta_i$ unrestricted.

Lemma 2. Under the conditions laid out in Section 2, as $N \to \infty$ with $\theta > 1/2$,

(a) $A_{1N}^* = \sigma^2(\lambda_{1N}^* + \tau_{N}\gamma_{1N}^*) + o_p(1)$,

(b) $A_{2N}^* - \sqrt{N}\sigma^2\theta_N \to_d \sigma^2 \left( \lim_{N \to \infty} (\omega_N + \tau_{N}\gamma_{0N}^* - \theta_N^2) \right)^{1/2} N(0,1)$,

(c) $B_N^* = \sigma^2(\lambda_{0N}^* + \tau_{N}\gamma_{0N}^*) + o_p(1)$,
where

\[
\begin{align*}
\lambda_{pN}^* &= \sum_{j=0}^{\infty} \frac{(2^{j-1}(j-1) + 1)}{2^{j-1}(j+3)} \phi_{jN}\mu_{j+p} \\
\gamma_{pN}^* &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (j+2)(k+2)\phi_k\phi_j \lambda_{Nj+k}^l + \\
&- \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(k+2)(l+2)}{2^k} \phi_{kN}\phi_{jN}\lambda_{Nj+k+l}^l, \\
\theta_N &= -\sum_{j=0}^{\infty} \frac{\phi_{jN}\mu_j}{2^j}, \\
\sigma^2_N &= \sum_{j=0}^{\infty} \frac{(2^j(j+3) - 2(2^{j+1} - j))(j+4) + 4(2^{j+3} - j - 5)}{2^j(j+3)(j+4)} \phi_{jN}\mu_j.
\end{align*}
\]

An important difference when compared to the results reported in Lemma 1 is that while \(A_{2N}^*\) is (asymptotically) mean zero \(A_{2N}^*\) is not. This is due to the demeaning and the fact that \(\bar{y}_t\) is correlated with \(\epsilon_{i,t}\), which calls for some kind of bias correction. As Moon and Perron (2008) point out, there are essentially two ways in which such a correction might be performed; either we bias-correct the numerator only, or we bias-correct the entire test statistic. Let us therefore consider as an example the following test statistic with a corrected numerator:

\[
t_0^* = \frac{A_N - \sqrt{N}\hat{o}^2 \theta_N}{\hat{o}/\sqrt{B_N}} = \sqrt{N}\hat{o}_N \frac{A_{1N}^*}{\hat{o}/\sqrt{B_N}} + \frac{A_{2N}^* - \sqrt{N}\hat{o}^2 \theta_N}{\hat{o}/\sqrt{B_N}} + o_p(1), \tag{9}
\]

where the second equality follows from the consistency of \(\hat{o}^2\) and assuming \(\eta + \theta \geq 1/4\) (as in (5)). According to Lemma 2,

\[
\frac{A_{1N}^*}{\hat{o}/\sqrt{B_N}} = \frac{\lambda_{0N}^* + \tau_N \gamma_{0N}^*}{\lambda_{0N}^* + \tau_N \gamma_{0N}^*} + o_p(1), \tag{10}
\]

\[
\frac{A_{2N}^* - \sqrt{N}\hat{o}^2 \theta_N}{\hat{o}/\sqrt{B_N}} \rightarrow_d \left( \lim_{N \rightarrow \infty} \frac{\omega_N + \tau_N \gamma_{0N}^* - \theta_N^2}{\lambda_{0N}^* + \tau_N \gamma_{0N}^*} \right)^{1/2} N(0,1) \tag{11}
\]

as \(N \rightarrow \infty\) with \(\theta > 1/2\). The problem with \(t_0^*\) is that it is not really feasible, as both the mean in (10) and the variance in (11) depend on unknowns. However, note that under the null, \(\omega_N = 1/3, \theta_N = -1/2, \lambda_{0N}^* = 1/6\) and \(\gamma_{pN}^* = 0\). Thus, letting

\[
t^* = \frac{\sqrt{2}(A_N^* + \sqrt{N}\hat{o}^2/2)}{\hat{o}/\sqrt{B_N}},
\]

we can show that (again under the null),

\[
\left( \frac{\omega_N + \tau_N \gamma_{0N}^* - \theta_N^2}{\lambda_{0N}^* + \tau_N \gamma_{0N}^*} \right)^{-1/2} \frac{A_N^* - \sqrt{N}\hat{o}^2 \theta_N}{\hat{o}/\sqrt{B_N}} = t^* + o_p(1) \rightarrow_d N(0,1) \tag{12}
\]

13
as \( N \to \infty \) with \( \theta > 1/2 \), suggesting that \( t^* \) can be used as a test statistic. Thus, as in the case
when \( \beta_i = 0 \), the asymptotic null distribution of \( t^* \) is independent of the initialization. The
local power function of this test statistic, which is identically the Moon and Perron (2008) \( t^\# \)
statistic, can be deduced from
\[
t^* = \frac{\sqrt{2N}(\alpha_N\lambda^*_1 + \tau_N\gamma^*_1) + \theta_N + \frac{1}{2}}{\sigma/\sqrt{B_N^*}} + \frac{\sqrt{2}(A_{2N}^* - \sqrt{N}\sigma^2\theta_N)}{\sigma/\sqrt{B_N^*}} + o_p(1),
\]
where the second term on the right-hand side is as in (11) (except for the multiplication by \( \sqrt{2} \)). As for the first term, which determines the mean of the test statistic, by using Taylor
expansion and the fact that
\[
a_N(\lambda^*_1 + \tau_N\gamma^*_1) + \theta_N + \frac{1}{2} = \frac{\alpha^2_N\mu_2}{24} + \frac{\alpha^3_N\mu_3}{24}(1 + 2\tau_N) + O_p(\alpha^4_N),
\]
we obtain
\[
\frac{\sqrt{2N}(\alpha_N\lambda^*_1 + \tau_N\gamma^*_1) + \theta_N + \frac{1}{2}}{\sigma/\sqrt{B_N^*}} = \frac{\sqrt{12N}\alpha^2_N\mu_2}{24} + O_p(\sqrt{N}\alpha^3_N),
\]
with the \( O_p(\sqrt{N}\alpha^3_N) \) remainder capturing the dependence on both \( \tau_N \) and higher order moments of \( c_i \).
Hence, unlike the situation when \( \beta_i = 0 \), provided that \( O_p(\sqrt{N}\alpha^3_N) = o_p(1) \),
the mean of \( t^* \) is asymptotically independent of \( \tau_N \). Because the same is true for the variance
in (11), we have that when \( \kappa \in [0,1] \) the asymptotic distribution of \( t^* \) is completely independent of the initialization. Thus, not only is size independent of the initialization, but provided that \( \kappa \in [0,1] \), so is power.

When \( \kappa > 1 \) the mean and variance of the statistic simplify to
\[
\frac{\alpha_N(\lambda^*_1 + \tau_N\gamma^*_1) + \theta_N + 1/2}{\sqrt{\lambda^*_0 + \tau_N\gamma^*_0}} = \alpha_N\sqrt{\tau_N}\left[\frac{\gamma^*_1}{\gamma^*_0} + O\left(\frac{1}{\tau_N}\right)\right]
\]
\[
= \frac{\alpha^2_N\sqrt{\tau_N}\mu_3}{\sqrt{24}\mu_2} + O\left(\sqrt{\tau_N}\alpha^3_N\right) + O\left(\frac{\alpha_N}{\sqrt{\tau_N}}\right),
\]
\[
\frac{\omega_N + \tau_N\gamma^*_1 - \theta^2_N}{\lambda^*_0 + \tau_N\gamma^*_0} = 1 + O\left(\frac{1}{\tau_N}\right).
\]
Thus, just as in the case when \( \beta_i = 0 \), the initial value has a dominating effect. However, since \( O(1/\tau_N) = o(1) \), the effect works only through the mean of the test statistic. Another implication of (16) is that the appropriate test statistic to use when \( \kappa > 1 \) is no longer given by \( t^* \) but rather by \( t^*/\sqrt{2} \).

Remarks.
10. The fact that when $\kappa \in [0, 1]$ the asymptotic distribution of $t^*$ does not depend on the initialization stands in stark contrast to previous results. In fact, the power of all unit root tests proposed in the literature so far, including the Im et al. (2003) panel test considered by Harris et al. (2010), depend on the initial condition (even when fixed effects are allowed for), and usually in a negative way.

11. For power to be non-negligible and non-increasing when $\beta_i$ is unrestricted and $\kappa \in [0, 1]$, according to (14), we need $\sqrt{N}\alpha_N^2 = 1$, which is satisfied for $\eta = 1/4$ but not for $\eta = 1/2$ (as in the case when $\beta_i = 0$). The demeaning therefore has an order effect on the neighborhood around unity for which power is negligible, which is in agreement with the results reported by Moon and Perron (2008, Theorem 4.1) for the case when $\kappa = 0$. The condition for non-negligible power when $\kappa > 1$ is given by $\sqrt{N\tau_N\alpha_N^2} = N^\theta(\kappa - 1/2 - 2\eta + 1/2) = 1$ (see (15)), which is satisfied if $\eta = (\theta(\kappa - 1) + 1)/4 > 1/4$. This again illustrates the power increasing potential of the initial value when in the infinite distant past.\footnote{The minimum requirement for power when $\eta = 1/2$ is that $\theta \geq 3/(\kappa - 1)$. Thus, if $\kappa = 2$, then we need $\theta \geq 3$, which is obviously quite restrictive, even in the typical macroeconomic application with $T \gg N$.}

12. The fact that $t^*$ has no power in $1/\sqrt{NT}$-neighborhoods of the null when $\kappa \in [0, 1]$ is due to the fact that we have replaced $\theta_N$ by $-1/2$, which is only appropriate under $H_0$. Knowledge of $\theta_N$ makes the otherwise infeasible test statistic in (9) operational, and this statistic has power also when $\eta = 1/2$.\footnote{Of course, if all the moments of $c_i$ were known there would be no point in testing for a unit root in the first place.}

13. By plugging in $\eta = 1/4$ and $\kappa \in [0, 1]$, and then taking the limit as $N \to \infty$, we obtain $t^* \to_d -\mu_2/4\sqrt{3} + N(0, 1)$, which is the same result as the one reported by Moon and Perron (2008, Theorem 4.2) for the case when $\kappa = 0$. This first-order theory leads to very simple predictions, as power can only stem from $\mu_2$. However, as (14) makes clear, this need not be the case. In particular, given the relatively slow rate of convergence in this case, in small-samples there is a potential offsetting effect as higher moments (as captured by the $\sqrt{N\tau_N\alpha_N^3}$ remainder term), while asymptotically negligible, may come into play. Similarly, although asymptotically there should be no dependence on $\tau_N$, in finite samples the initialization can have an effect on power.

14. The analysis of the case when the bias-correction is done to the entire test statistic is
entirely analogous to the one presented above. To illustrate this, consider the following infeasible test statistic:

\[
\frac{A^*_N}{\hat{\sigma} \sqrt{B^*_N}} - \frac{\sqrt{N} \theta_N}{\sqrt{\lambda^*_{0N} + \tau_N \gamma_{0N}}} = \frac{A^*_N - \sqrt{N} \hat{\sigma}^2 \theta_N}{\hat{\sigma} \sqrt{B^*_N}} + \sqrt{N} \theta_N \left[ \frac{\hat{\sigma}}{\sqrt{B^*_N}} - \frac{1}{\sqrt{\lambda^*_{0N} + \tau_N \gamma_{0N}}} \right],
\]

where the first term on the right-hand side is as in (10), while the second can be expanded as

\[
\sqrt{N} \theta_N \left[ \frac{\hat{\sigma}}{\sqrt{B^*_N}} - \frac{1}{\sqrt{\lambda^*_{0N} + \tau_N \gamma_{0N}}} \right] = -\frac{\theta_N}{2(\lambda^*_{0N} + \tau_N \gamma_{0N})^{3/2}} \sqrt{N} \left[ \frac{B^*_N}{\hat{\sigma}^2} - (\lambda^*_{0N} + \tau_N \gamma_{0N}) \right] + o_p(1).
\]

Since this term is mean zero, it is clear that power, and hence also the dependence on \(\tau_N\), will be driven by the first term.

15. In contrast to Lemma 1, Lemma 2 assumes that \(\theta > 1/2\), which implies \(\sqrt{N}/T \to 0\) as \(N, T \to \infty\). The reason for this requirement is the fixed effects, whose elimination induces an estimation error in \(T\), which is then aggravated when pooling across \(N\). The condition that \(\sqrt{N}/T \to 0\) prevents this error from having a dominating effect.

4 Simulations

A small-scale simulation study was conducted to assess the accuracy of our theoretical results in small samples. The data generating process is given by (1)–(3), where \(\epsilon_{i,t} \sim N(0, 1)\), \(\beta_i = 0, \delta = 1\) and \(c_i \sim U(a, b)\), suggesting that \(\mu_m = \sum_{s=0}^{m} \delta^s b^{m-s}/(m + 1)\). The data are generated for 3,000 panels with \(\theta = 1\). To also ensure that \(\tau_N \alpha_N \leq 1\) in the case when \(\kappa > 1\) (see remark 2), we set \(\kappa \leq \eta + 1\), where \(\eta = 1/4\) when fixed are allowed for and \(\eta = 1/2\) when not.

\(t\) and \(t^*\) are implemented with \(\hat{\sigma}^2\) set equal to \(\hat{\sigma}^2 = \sum_{i=1}^{N} \sum_{t=2}^{T} (\Delta y_{i,t})^2 / NT\). The results are compared with the infinite-order power functions derived in Section 3 and also with the corresponding first-order power functions based on assuming \(T_0 = 0\) (as summarized in Moon et al., 2007). We also report the results from the infeasible test statistic in the fixed effects.
model, \( t_0^* \), and its power function. All tests are carried out at the 5% level, and the infinite-order power functions are truncated such that they only include the first 100 moments of \( c_i \). All tests are double-sided.\(^9\)

The results reported in Tables 1–3 are generally in agreement with theory and can be summarized as follows:

- The size accuracy of the tests is generally very good. There are some distortions but these diminish as \( N \) increases. As expected, the results are not sensitive to the initialization (as measured by \( \kappa \)).

- Power is generally very close what is predicted by the (truncated) infinite-order power functions of Section 3. When \( \kappa = 0 \), the predictions based on the first-order power functions are also quite close, but only when \( a \) and \( b \) are relatively close to zero, and the precision deteriorates as \( \kappa \) and the deviation from the null increases. In fact, the first-order theory is way off target in most cases, especially in the fixed effects model. To take an extreme example, consider the case when \( a = b = -6 \) and \( \kappa = 0, 1 \), in which predicted power based on the first-order theory is about 100 times as large as actual power. The fact that the bias is mainly driven by \((a, b)\) suggests that the poor accuracy is not due to the initialization but rather to the error coming from the first-order approximation.

- As expected, when \( \beta_i = 0 \) and \( \eta = 1/2 \) power is driven mainly by \( \mu_1 \). However, there is also a second-order effect working through variance of \( c_i \). In particular, while the first-order theory completely misses this, both the empirical and infinite-order theoretical power seem to be decreasing in \(|a - b|\).

- When \( \beta_i = 0 \) and \( \kappa = 3/2 \), since \( \theta = 1 \) and \( \eta = 1/2 \), we have \( \sqrt{N}r_N \alpha_N = N^{1/4} \), suggesting that in this case power should be increasing in \( N \). The results reported in Table 1 are quite suggestive of this.

- In the fixed effects model, since \( \eta = 1/4 \) in the simulations, the power of \( t_0^* \) should go to one with the sample size, and this just what we see in the table. In fact, there are only a few occasions when power less than 100%. The power of \( t^* \) is, on the other hand, very poor and it is only rarely that power actually raises above the nominal 5% level.

\(^9\)The one-sided rejection frequencies are available from the author upon request.
We also see that, as alluded in remark 4 of Section 3.2, if $\kappa \in [0, 1]$, while asymptotically there should be no dependence on $\kappa$, there is some variation in the results depending on whether $\kappa = 0$ or $\kappa = 1$. Again, while the existing theory completely misses this, our asymptotic theory is able to capture the effect of $\kappa$.

5 Conclusion

The results obtained here are interesting in their own right but also because of the implications they have for applied work. The fact that with fixed effects the pooled unit root $t$-statistic considered here is asymptotically invariant with respect to distant past initialization is, for example, extremely useful. Motivated by the otherwise so common relationship between power and size of the initial value, much effort has recently gone into the development of “robust” testing strategies (see, for example, Elliott and Müller, 2006; Harvey et al., 2009). Unfortunately, such strategies cannot remove the effect of the initial value (see Elliott and Müller, 2006, for a formal proof), and therefore only provide a partial solution to the problem. They can also be quite difficult to implement, and no one strategy seem to dominate the others (see Harvey and Leybourne, 2006). Our results suggest that with panel data there is no need for robust testing strategies. In fact, since size is unaffected and power is either flat or increasing in the extent of the initialization, there seem to be no immediate cause for concern. In an essence, our results give some credence to the common empirical approach of simply ignoring the initial value.

Several interesting extensions of the current work come to mind. Firstly, although the introduction of incidental trends is unlikely to affect of the initialization, the current analysis can be extended to the case with detrended data. Secondly, and perhaps more importantly, while in the case of fixed effects the current paper focuses on the use of OLS demeaned data, one could also consider generalized least squares (GLS) demeaning (see Elliott et al., 1996). This avenue is currently being pursued by the author in a separate work.
References


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

Appendix: Proofs

Proof of Lemma 1.

Consider first the case when \( \kappa \in [0, 1] \). We begin with (a). Define \( y_{i,0} = \sum_{s=0}^{T_0-2} \rho_s^i \epsilon_{i,-s} \) such that

\[ y_{i,t} = \rho_t^i y_{i,0} + \sum_{s=1}^t \rho_t^{i-s} \epsilon_{i,s}. \]

Hence, by further defining \( r_{i,t} = \exp(c_i\alpha_i N \tau t / T) \) and \( r^0_{i,t} = \exp(c_i\alpha_i \tau N t / T_0) \), we have

\[ \frac{1}{\sqrt{T}} y_{i,t} = \frac{1}{\sqrt{T}} \rho_t^i y_{i,0} + \frac{1}{\sqrt{T}} \sum_{s=1}^t \rho_t^{i-s} \epsilon_{i,s} = r_{i,t} \sqrt{\tau N} x^0_{i,T_0} + x_{i,t}, \]  

where, by virtue of the serial independence of \( \epsilon_{i,t} \), the processes

\[ x^0_{i,t} = \frac{1}{\sqrt{T}} \sum_{s=0}^t r^0_{i,s} \epsilon_{i,-s}, \]

\[ x_{i,t} = \frac{1}{\sqrt{T}} \sum_{s=1}^t r_{i,t-s} \epsilon_{i,s} \]

are independent of each other. It follows that

\[
E\left( \frac{1}{T^2} \sum_{t=2}^T c_i y_{i,t-1}^2 | c_i \right) = E\left( \frac{1}{T} \sum_{t=2}^T c_i (r_{i,t-1} \sqrt{\tau N} x^0_{i,T_0} + x_{i,t-1})^2 | c_i \right) = \tau N E[c_i (x^0_{i,T_0})^2 | c_i] \frac{1}{T} \sum_{t=2}^T r^2_{i,t-1} + \frac{1}{T} \sum_{t=2}^T E(c_i x^2_{i,t-1} | c_i). \]  

(A2)

Assume that \( \tau_N \to \tau \) and \( \alpha_N \to \alpha \) as \( N \to \infty \), where \( \tau, \alpha \in \{0, 1\} \) (which is without loss of generality, as we have already assumed that \( \kappa \in [0, 1] \)). Hence, since conditional on \( c_i \), \( r_{i,t} \) is purely deterministic, \( r_{i,t} | c_i \to r_i(w) = \exp(c_i \alpha w) \) as \( N \to \infty \) (implying \( T \to \infty \)), where \( t = \lfloor wT \rfloor, w \in [0, 1] \) and \( \lfloor x \rfloor \) denotes the integer part of \( x \). This implies, with \( s = \lfloor vT \rfloor \) and \( v \leq w \),

\[
E(x_{i,t} x_{i,s} | c_i) = \frac{1}{T} \sum_{m=1}^s \sum_{n=1}^s r_{i,t-m} r_{i,s-n} E(\epsilon_{i,m} \epsilon_{i,n} | c_i) = \sigma^2 \left( \frac{1}{T} \sum_{n=1}^s r_{i,t-n} r_{i,s-n} \right) \to \sigma^2 \int_{u=0}^v r_i(w-u) r_i(v-u) du = 2 \sigma^2 \int_{u=0}^v r_i(w+v-2u) du \to \sigma^2 \frac{1}{2c_i \alpha} (r_i(w+v) - r_i(w-v)). \]  

(A3)

We similarly have \( r^0_{i,t} | c_i \to r^0_{i,t}(w) = \exp(c_i \alpha \tau w) \), suggesting that

\[
E(x^0_{i,t} x^0_{i,s} | c_i) \to \sigma^2 \frac{1}{2c_i \alpha \tau} (r^0_{i}(w+v) - r^0_{i}(w-v)). \]  

(A4)
Therefore, since \( r_i(0) = r_i^0(0) = 1 \),
\[
E[(x_{i,t}^0)^2|c_i] \rightarrow \sigma^2 \frac{1}{2c_i\alpha \tau} (r_i^0(2w) - 1),
\]
\[
E(x_{i,t}^2|c_i) \rightarrow \sigma^2 \frac{1}{2c_i\alpha} (r_i(2w) - 1)
\]
as \( N \rightarrow \infty \). By using this, \( j! = (j - 1)! j \) and then Taylor expansion of the type \( \exp(x) = \sum_{j=0}^{\infty} x^j / j! \), we obtain
\[
E[(x_{i,T_0}^0)^2|c_i] \rightarrow \sigma^2 \frac{1}{2c_i\alpha \tau} (r_i^0(2) - 1) = \sigma^2 \frac{1}{2c_i\alpha} \left( \sum_{j=0}^{\infty} \frac{(2\alpha\tau)^j}{j!} \right) - 1 = \sigma^2 \sum_{j=0}^{\infty} \frac{(2\alpha\tau)^j}{(j+1)!}
\]
\[
= \sigma^2 \sum_{j=0}^{\infty} (j+2) \phi_j (\tau c_i)^j.
\]
Similarly, since
\[
\frac{1}{T} \sum_{t=2}^{T} E(x_{i,t}^2|c_i) \rightarrow \sigma^2 \frac{1}{2c_i\alpha} \int_0^1 (r_i(2w) - 1) dw = \sigma^2 \frac{1}{2c_i\alpha} \left( \frac{1}{2c_i\alpha} (r_i(2) - 1) - 1 \right)
\]
\[
= \sigma^2 \sum_{j=0}^{\infty} \frac{(2c_i\alpha)^j}{(j+2)!} = \sigma^2 \sum_{j=0}^{\infty} \phi_j c_i^j
\]
as \( N \rightarrow \infty \), we can show that
\[
\frac{1}{T} \sum_{t=2}^{T} E(c_i x_{i,t}^2) = \frac{1}{T} \sum_{t=2}^{T} E[c_i E(x_{i,t}^2|c_i)] \rightarrow \sigma^2 \sum_{j=0}^{\infty} \frac{(2\alpha)^j}{(j+2)!} \phi_{j+1} = \sigma^2 \sum_{j=0}^{\infty} \phi_j \mu_{j+1}.
\]
Also, from \( r_i(w)^p = r_i(pw) \) for \( p > 0 \),
\[
\frac{1}{T} \sum_{t=2}^{T} r_{i,t}^2|c_i \rightarrow \int_0^1 r_i(2w) dw = \frac{1}{2c_i\alpha} (r_i(2) - 1) = \sum_{j=0}^{\infty} \frac{(2\alpha c_i)^j}{(j+1)!} = \sum_{j=0}^{\infty} (j+2) \phi_j c_i^j,
\]
and therefore
\[
E \left( \frac{1}{T^2} \sum_{t=2}^{T} c_i y_{i,t-1}^2 \right) = \frac{1}{T^2} \sum_{t=2}^{T} E[c_i x_{i,t-1}^2|c_i] = \tau N E \left( E[c_i(x_{i,T_0}^0)^2|c_i] \frac{1}{T^2} \sum_{t=2}^{T} r_{i,t-1}^2 \right)
\]
\[
\rightarrow \sigma^2 \sum_{j=0}^{\infty} \phi_j \mu_{j+1} + \sigma^2 \tau \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (j+2)(k+2) \phi_j \phi_k \mu_{j+k+1}
\]
\[
= \sigma^2 (\lambda_1 + \tau \gamma_1).
\]
We now verify conditions (i) and (ii) of the law of large numbers of Phillips and Moon (1999, Corollary 1), as is done in, for example, Moon and Phillips (2000, page 975). In their notation, we have \( Q_{i,T} = \sum_{t=2}^{T} c_i y_{i,t-1}^2 / T^2 \), which is iid across \( i \). The following convergence results hold: \( x_{i,T_0}^0 \rightarrow_w \sigma_i^0(w) \) and \( x_{i,t} \rightarrow_w \sigma_i(w) \), where \( \rightarrow_w \) signifies weak convergence, \( t_0 = [wT_0] \), \( t = [wT] \), and \( J_i(w) = \int_{v=0}^{w} r_i(w - v) dW_i(v) \) is a diffusion process that is independent
of the reverse diffusion \( f_0(w) = \int_{v=0}^{w} f_0(v) dW_0(v) \). \( W_0(v) \) and \( W_0'(v) \) are independent standard Brownian motions. Making use of these results, letting \( K_i(w) = r_i(w)\sqrt{T}f_0(1) + J_i(w) \), we have

\[
Q_{i,T} = \frac{1}{T} \sum_{t=2}^{T} c_i(r_{i,t-1}\sqrt{T}x_{i,t_0}^0 + x_{i,t-1})^2 \rightarrow w \sigma^2 \int_{w=0}^{1} c_i K_i(w)^2 dw = Q_i
\]
as \( N \rightarrow \infty \). Moreover, it is clear from the above that \( E(Q_i) = \sigma^2(\lambda_1 + \tau \gamma_1) \). Therefore, since \( Q_{i,T} \rightarrow w Q_i \) and \( E(Q_{i,T}) \rightarrow E(Q_i) \), we have that \( |Q_{i,T}| \) is uniformly integrable (see Moon and Phillips, 2000, page 971). This establishes condition (i). Moreover, since the scaling coefficient ("C_i" in Phillips and Moon, 1999) is just one, (ii) is automatically satisfied. We can therefore show that

\[
A_{1N} = \frac{1}{N} \sum_{i=1}^{N} Q_{i,T} \rightarrow p \sigma^2(\lambda_1 + \tau \gamma_1)
\]
as \( N \rightarrow \infty \), where \( \rightarrow p \) signifies convergence in probability.

In order to prove (b) we use the same steps as in Moon and Phillips (2000, page 994) to verify that \( Q_{i,T} = \sum_{t=2}^{T} y_{i,t-1} e_{i,t}/T \) satisfies conditions (i)–(iv) of the central limit theorem of Phillips and Moon (1999, Theorem 2). We begin by noting that \( Q_{i,T} \) is iid, and, denoting by \( F_i \) the sigma-field generated by \( \{e_{i,t}\}_{s=-T_0+2} \), it is further clear that

\[
E(Q_{i,T}) = \sqrt{\tau N}E \left( x_{i,t_0}^0 \frac{1}{\sqrt{T}} \sum_{t=2}^{T} r_{i,t-1} E(e_{i,t}|F_{i,t-1}) \right) + E \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} x_{i,t-1} E(e_{i,t}|F_{i,t-1}) \right) = 0.
\]

As for the variance of \( Q_{i,T} \), we have

\[
E(Q_{i,T}^2|c_i) = E \left( \left( \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} e_{i,t} \right)^2 |c_i \right)
\]

\[
= E \left( \frac{1}{T} \sum_{t=2}^{T} x_{i,t-1}^2 E(e_{i,t}^2|F_{i,t-1}) |c_i \right) + \tau N E \left( (x_{i,t_0}^0)^2 \frac{1}{T} \sum_{t=2}^{T} r_{i,t-1}^2 E(e_{i,t}^2|F_{i,t-1}) |c_i \right)
\]

\[
= \frac{1}{T} \sum_{t=2}^{T} E(x_{i,t-1}^2 |c_i) \sigma^2 + \tau N \sum_{t=2}^{T} r_{i,t-1}^2 \sigma^2
\]

\[
\rightarrow \sigma^4 \sum_{j=0}^{\infty} \phi_j c_j^i + \sigma^4 \tau \sum_{j=0}^{\infty} (j+2)(k+2) \tau^j \phi_j \phi_k c_j^i k
\]
as \( N \rightarrow \infty \), giving

\[
E(Q_{i,T}) \rightarrow \sigma^4 \sum_{j=0}^{\infty} \phi_j \mu_j + \sigma^4 \tau \sum_{j=0}^{\infty} (j+2)(k+2) \tau^j \phi_j \phi_k \mu_{j+k} = \sigma^4(\lambda_0 + \tau \gamma_0),
\]

(A7)
which is finite. Moreover,

$$Q_{i,T} = \frac{1}{T} \sum_{t=2}^{T} y_{i,t-1} \epsilon_{i,t} \to_w \sigma^2 \int_{w=0}^{1} K_i(w) dW_i(w) = Q_i$$

as $N \to \infty$, and it is not difficult to verify that $E(Q_i^2) = \sigma^4(\lambda_0 + \tau \gamma_0)$. Together $Q_{i,T} \to_w Q_i$ and $E(Q_{i,T}^2) \to E(Q_i^2)$ imply that conditions (i), (ii) and (iv) are satisfied. Condition (iv) follows from noting that, by the continuous mapping theorem, $Q_{i,T}^2 \to_w Q_i^2$. It follows that

$$A_{2N} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Q_{i,T} \to_d \sigma^2 \sqrt{\lambda_0 + \tau \gamma_0} N(0,1) \quad (A8)$$

as $N \to \infty$.

Finally, consider (c). $\sum_{t=2}^{T} y_{i,t-1}^2 / T^2$ satisfies the conditions of Corollary 1 of Phillips and Moon (1999) (which can be verified in the same way as in the proof of (a)). Therefore,

$$B_N = \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^2 \to_p \sigma^2(\lambda_0 + \tau \gamma_0) \quad (A9)$$

as $N \to \infty$. This completes the proof for the case when $\kappa \in [0,1]$. The results for the case when $\kappa > 1$ are the same, provided that $\alpha N \tau N \leq \frac{1}{2}$, thus ensuring that the limit of $r_{i,t}(w) = \exp(c_i \alpha N \tau N t / T_0)$ exists and that it admits to the Taylor expansion $\exp(x) = \sum_{j=0}^{\infty} x^j / j!$.

**Proof of Lemma 2.**

As in the proof of Lemma 1 we assume that $\kappa \in [0,1]$. The results for the case when $\kappa > 1$ are the same if we assume that $\alpha N \tau N \leq 1$ (see the discussion following (A9) above).

Consider (a). By using

$$\frac{1}{\sqrt{T}} (y_{i,t} - \bar{y}_i) = (r_{i,t} - \bar{r}_i) \sqrt{\tau N} x_{i,t}^0 + x_{i,t} - \bar{x}_i,$$

and the independence of $x_{i,t}^0$ and $x_{i,t}$, we obtain

$$E \left( \frac{1}{T^2} \sum_{t=2}^{T} c_i (y_{i,t-1} - \bar{y}_{i-1})^2 | c_i \right) = \tau_N E[c_i (x_{i,T_0})^2 | c_i] \frac{1}{T} \sum_{t=2}^{T} (r_{i,t-1} - \bar{r}_{i-1})^2 + \frac{1}{T} \sum_{t=2}^{T} E[c_i (x_{i,t-1} - \bar{x}_{i-1})^2 | c_i], \quad (A10)$$
where
\[
\frac{1}{T} \sum_{t=2}^{T} (r_{i,t} - \bar{r}_i)^2 |c_i| = \frac{1}{T} \sum_{i=2}^{T} \sum_{j=0}^{\infty} \frac{(j + 2)}{2j} (k + 2) \phi_j c_{i}^{j+k},
\]
\[
\frac{1}{T} \sum_{t=2}^{T} E[(x_{i,t-1} - \bar{x}_{i-1})^2 |c_i] = \frac{1}{T} \sum_{t=2}^{T} E(x_{i,t-1}^2 |c_i) - E(\bar{x}_{i-1}^2 |c_i)
\]
\[
\rightarrow \sum_{j=0}^{\infty} \frac{(j + 2)}{2j} \phi_j c_{i}^{j+k},
\]

as \( N \to \infty \). The former result uses that
\[
\bar{r}_i |c_i| = \frac{1}{T} \sum_{t=2}^{T} r_{i,t} |c_i| \rightarrow \int_{w=0}^{1} r_i(w) dw = \frac{1}{c_i} (r_i(1) - 1) = \sum_{j=0}^{\infty} \frac{\sigma_j}{(j+1)!} c_{i}^{j} = \sum_{j=0}^{\infty} \frac{(j + 2)}{2j} \phi_j c_{i}^{j},
\]
whereas the latter uses that, because of symmetry,
\[
E(\bar{x}_{i-1}^2 |c_i) = \frac{1}{T} \sum_{t=2}^{T} \sum_{s=2}^{T} E(x_{i,t-1}x_{i,s-1} |c_i)
\]
\[
\rightarrow \sum_{j=0}^{\infty} \frac{(2j-1)(j-1)}{4(j+3)!} \phi_j c_{i}^{j},
\]

By adding these results,
\[
E \left( \frac{1}{T^2} \sum_{t=2}^{T} c_i (y_{i,t-1} - \bar{y}_{i-1})^2 |c_i \right)
\]
\[
\rightarrow \sigma^2 \sum_{j=0}^{\infty} \frac{(2j-1)(j-1)}{4(j+3)!} \phi_j c_{i}^{j+1}
\]
\[
+ \tau \sigma^2 \sum_{j=0}^{\infty} \frac{(j + 2)}{2j} \phi_j T c_{i}^{j+1} \left( \sum_{k=0}^{\infty} \frac{(k + 2)}{2k} \phi_k c_{i}^{k} - \sum_{k=0}^{\infty} \frac{(k + 2)}{2l} \phi_k \phi_l c_{i}^{k+l} \right)
\]
25
as \( N \to \infty \), and therefore,
\[
E \left( \frac{1}{T^2} \sum_{t=2}^{T} c_i(y_{i,t-1} - \bar{y}_{i,-1})^2 \right) \to \sigma^2 \sum_{j=0}^{\infty} \frac{(2^j - 1) + 1}{2^{j-1}(j + 3)} \phi_j \mu_{j+1}
\]
\[
+ \tau \sigma^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (j + 2)(k + 2) \phi_k \phi_j \tau^j \mu_{j+k+1}
\]
\[
- \tau \sigma^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (j + 2)(l + 2) \frac{(k + 2)}{2^k} \phi_k \phi_j \tau^j \mu_{j+k-l+1}
\]
\[
= \sigma^2 (\lambda_1^* + \tau \gamma_1^*). \hspace{1cm} (A11)
\]

By following the same steps used in the proof of Lemma 1 (a) it is possible to show that
\[
\sum_{t=2}^{T} c_i(y_{i,t-1} - \bar{y}_{i,-1})^2 / T^2
\]
satisfies the conditions of Corollary 1 of Phillips and Moon (1999).

Hence,
\[
A_{IN}^* = \frac{1}{NT^2} \sum_{t=1}^{N} \sum_{i=2}^{T} c_i(y_{i,t-1} - \bar{y}_{i,-1})^2 \to \sigma^2 (\lambda_1^* + \tau \gamma_1^*) \hspace{1cm} (A12)
\]
as \( N \to \infty \).

Consider (b). We have
\[
\frac{1}{T} \sum_{t=2}^{T} (y_{i,t-1} - \bar{y}_{i,-1}) \epsilon_{i,t} = \sqrt{T} \alpha_{i,T_0} \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (r_{i,t-1} - \bar{r}_{i,-1}) \epsilon_{i,t} + \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i,-1}) \epsilon_{i,t}, \hspace{1cm} (A13)
\]
where the first term on the right-hand side is clearly mean zero, while, for \( s = \lfloor vT \rfloor \), \( t = \lfloor wT \rfloor \) and \( v > w \),
\[
\frac{1}{\sqrt{T}} \sum_{t=2}^{T} E[(x_{i,t-1} - \bar{x}_{i,-1}) \epsilon_{i,t} | c_i]
\]
\[
= - \frac{1}{\sqrt{T}} \sum_{t=2}^{T} E(\bar{x}_{i,-1} \epsilon_{i,t} | c_i) = - \frac{1}{T^{3/2}} \sum_{t=2}^{T} \sum_{s=2}^{T} E(x_{i,s-1} \epsilon_{i,t} | c_i)
\]
\[
\to - \sigma^2 \int_{v=0}^{1} \int_{w=0}^{v} r_i(v - w) dw dv = - \sigma^2 \frac{1}{c_i \alpha} \int_{v=0}^{1} (r_i(v) - 1) dv
\]
\[
= - \sigma^2 \frac{1}{c_i \alpha} \left( \frac{1}{c_i \alpha} (r_i(1) - 1) - 1 \right) = - \sigma^2 \sum_{j=0}^{\infty} \frac{(c_i \alpha)^j}{(j + 2)} = - \sigma^2 \sum_{j=0}^{\infty} \frac{1}{2^j} \phi_j c_i
\]
as \( N \to \infty \), suggesting
\[
\frac{1}{\sqrt{T}} \sum_{t=2}^{T} E[(x_{i,t-1} - \bar{x}_{i,-1}) \epsilon_{i,t}] = \sigma^2 \theta + o(1), \hspace{1cm} (A14)
\]
where \( \theta = - \sum_{j=0}^{\infty} \phi_j \mu_j / 2^j \). The order of the error of approximation reported here is not the sharpest possible. In order to work out this order more exactly we make use of the following:
\[
\sup_{1 \leq t \leq T} \sup_{(t-1) / T \leq r \leq t / T} |(t/T)^k - r^k| = O \left( \frac{1}{T} \right),
\]
\[
\sup_{1 \leq t \leq T} \sup_{(t-1) / T \leq r \leq t / T} |\exp(t/T) - \exp(r)| = O \left( \frac{1}{T} \right)
\]
(see Moon and Phillips, 2000, page 992), from which it is possible to show that

\[
\frac{1}{\sqrt{T}} \sum_{t=2}^{T} E[(x_{i,t-1} - \bar{x}_{i,-1})\epsilon_{i,t}] = \sigma^2 \theta + O \left( \frac{1}{T} \right). \tag{A15}
\]

As for the variance of this term, we use

\[
\frac{1}{\sqrt{T}} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i,-1})\epsilon_{i,t} = \frac{1}{\sqrt{T}} \sum_{t=2}^{T} x_{i,t-1}\epsilon_{i,t} - \sqrt{T}\bar{x}_{i,-1}\epsilon_{i,t},
\]

suggesting that

\[
E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i,-1})\epsilon_{i,t} \right)^2 \right] = E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} x_{i,t-1}\epsilon_{i,t} \right)^2 \right] - 2E \left( \bar{x}_{i,-1}\epsilon_{i} \sum_{t=2}^{T} x_{i,t-1}\epsilon_{i,t} \right) + TE[(\bar{x}_{i,-1}\epsilon_{i})^2], \tag{A16}
\]

where the first term on the right-hand side is the same as in the proof of Lemma 1 (b).

Consider the second term. Clearly,

\[
E \left( \bar{x}_{i,-1}\epsilon_{i} \sum_{t=2}^{T} x_{i,t-1}\epsilon_{i,t} | c_i \right) = \frac{1}{T} \sum_{t=2}^{T} \sum_{s=2}^{T} E(x_{i,s-1}x_{i,t-1}\epsilon_{i,t}\epsilon_{i} | c_i)
\]

\[
= \frac{1}{T} \sum_{s=2}^{T} \sum_{t=2}^{T} E(x_{i,s-1}x_{i,t-1}\epsilon_{i,t}\epsilon_{i} | c_i) + \frac{1}{T} \sum_{s=2}^{T} \sum_{t=2}^{T} E(x_{i,s-1}x_{i,t-1}\epsilon_{i,t}\epsilon_{i} | c_i),
\]

where, with \( s = |vT| \), \( t = |wT| \) and \( w > v \),

\[
E(x_{i,s-1}x_{i,t-1}\epsilon_{i,t}\epsilon_{i} | c_i) = \frac{1}{T} \sum_{k=2}^{T} \sum_{m=2}^{T} E(x_{i,s-1}x_{i,t-1}\epsilon_{i,t}\epsilon_{i} | c_i) = \frac{1}{T} E(x_{i,s-1}x_{i,t-1}\epsilon_{i,t}^2 | c_i)
\]

\[
= \frac{1}{T} \sum_{k=2}^{T} \sum_{m=2}^{T} r_{i,t-1-k}r_{i,s-1-m}E(\epsilon_{i,k}\epsilon_{i,m}E(\epsilon_{i,t}^2 | F_{t-1}) | c_i)
\]

\[
= \sigma^4 \frac{1}{T^2} \sum_{k=2}^{T} \sum_{m=2}^{T} r_{i,t-1-k}r_{i,s-1-m}E(\epsilon_{i,k}\epsilon_{i,m} | c_i)
\]

\[
= \sigma^4 \frac{1}{T^2} \sum_{k=2}^{T} \sum_{m=2}^{T} r_{i,t-1-k}r_{i,s-1-m} \int_{u=0}^{w} r_i(w - u)r_i(v - u)du = \sigma^4 \int_{u=0}^{w} r_i(w + v - 2u)du
\]

\[
= \sigma^4 \frac{1}{2\epsilon_i\alpha}(r_i(w + v) - r_i(w - v))
\]
as \( N, T \to \infty \). It follows that

\[
\frac{1}{T} \sum_{s=2}^{T} \sum_{l=1}^{s-1} E(x_{i,s-1} x_{i,l-1} e_i \bar{e}_i | c_i) = \sigma^4 \frac{1}{T^3} \sum_{s=2}^{T} \sum_{l=2}^{s-1} \sum_{k=2}^{l-1} r_{i,s-1-k} r_{i,l-1-k} \\
\rightarrow \sigma^4 \frac{1}{2c_i \alpha} \int_{w=0}^{1} \int_{v=0}^{w} \int_{u=0}^{v} (r_i(w+v) - r_i(w-v)) dudvdw \\
= \sigma^4 \frac{1}{2(c_i \alpha)^2} \left( 1 - \frac{2}{c_i \alpha} (r_i(1) - 1) + \frac{1}{2c_i \alpha} (r_i(2) - 1) \right) \\
= \sigma^4 \sum_{j=0}^{\infty} \frac{(2j+1) - 1}{2(j+3)} \phi_j c_i^j.
\]

On the other hand, if \( w \leq v \),

\[
E(x_{i,s-1} x_{i,l-1} e_i \bar{e}_i | c_i) = \frac{1}{T} \sum_{k=2}^{l-1} E(x_{i,s-1} x_{i,l-1} e_i \bar{e}_i | c_i) + \frac{1}{T} E(x_{i,s-1} x_{i,l-1} e_i^2 | c_i) \\
+ \frac{1}{T} E(x_{i,l-1}^2 e_i | c_i),
\]

where the second term on the right-hand side is the same as before and

\[
\frac{1}{T^2} \sum_{s=2}^{T} \sum_{l=1}^{s-1} E(x_{i,l-1}^2 e_i^2 | c_i) = \sigma^2 \frac{1}{T^2} \sum_{s=2}^{T} E(x_{i,l-1}^2 | c_i) = O \left( \frac{1}{T} \right).
\]

As for the first term, we have

\[
\frac{1}{T} \sum_{k=2}^{l-1} E(x_{i,s-1} x_{i,l-1} e_i \bar{e}_i | c_i) = \frac{1}{T^2} \sum_{k=2}^{l-1} \sum_{m=2}^{l-1} \sum_{n=2}^{l-1} r_{i,s-1-m} r_{i,l-1-n} E(e_i \bar{e}_i e_i \bar{e}_i | c_i) \\
= \frac{1}{T^2} \sum_{k=2}^{l-1} \sum_{n=2}^{l-1} r_{i,s-1-n} r_{i,l-1-n} E(e_i \bar{e}_i e_i \bar{e}_i | c_i) \\
= \sigma^2 \frac{1}{T^2} \sum_{k=2}^{l-1} \sum_{n=2}^{l-1} r_{i,s-1-n} r_{i,l-1-n} E(e_i \bar{e}_i e_i \bar{e}_i | c_i) \\
= \sigma^4 \frac{1}{T^2} \sum_{k=2}^{l-1} \sum_{k=2}^{l-1} r_{i,s-1-k} r_{i,l-1-k},
\]

giving

\[
\frac{1}{T^2} \sum_{l=2}^{T} \sum_{s=1}^{l-1} E(x_{i,s-1} x_{i,l-1} e_i \bar{e}_i | c_i) = \sigma^4 \frac{1}{T^3} \sum_{l=2}^{T} \sum_{s=1}^{l-1} \sum_{k=2}^{l-1} r_{i,s-1-k} r_{i,l-1-k} \\
\rightarrow \sigma^4 \int_{w=0}^{1} \int_{v=0}^{w} \int_{u=0}^{v} r_i(v-w) r_i(w-u) dudvdw \\
= \sigma^4 \int_{w=0}^{1} \int_{v=w}^{v} \int_{u=0}^{v} r_i(v-u) dudvdw \\
= \sigma^4 \frac{1}{(c_i \alpha)^2} \left( 1 + r_i(1) - \frac{2}{c_i \alpha} (r_i(1) - 1) \right) \\
= \sigma^4 \sum_{j=0}^{\infty} \frac{(j+1)}{2(j+3)} \phi_j c_i^j.
\]
which in turn implies

\[
E \left( x_{i,t-1} \epsilon_1 \sum_{t=2}^{T} x_{i,t-1} \epsilon_i | c_i \right) \rightarrow \sigma^4 \sum_{j=0}^{\infty} \frac{2(j+1)}{2(j+3)} \phi_j c_i^j + 2 \sigma^4 \sum_{j=0}^{\infty} \frac{2(j+1)}{2(j+3)} \phi_j c_i^j
\]

\[
= \sigma^4 \sum_{j=0}^{\infty} \frac{2(j+2 + j-1)}{2(j+3)} \phi_j c_i^j
\]

(A17)

as \( N \to \infty \).

Next, consider \( TE[(x_{i,t-1} \epsilon_1)^2] \). We have

\[
TE(x_{i,t-1}x_{i,s-1} \epsilon_1^2 | c_i) = \frac{1}{T} \sum_{k=2}^{T-1} \sum_{m=2}^{T-1} E(x_{i,t-1}x_{i,s-1} \epsilon_i \epsilon_m | c_i)
\]

\[
+ 2 \frac{1}{T} \sum_{k=t}^{T} \sum_{m=2}^{T-1} E(x_{i,t-1}x_{i,s-1} \epsilon_i \epsilon_m | c_i) + \frac{1}{T} \sum_{k=1}^{T} \sum_{m=1}^{T} E(x_{i,t-1}x_{i,s-1} \epsilon_i \epsilon_m | c_i),
\]

where, for \( t > s \),

\[
\sum_{k=t}^{T} \sum_{m=1}^{T} E(x_{i,t-1}x_{i,s-1} \epsilon_i \epsilon_m | c_i) = \frac{1}{T} \sum_{k=t}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} E(e_i \epsilon_i \epsilon_m \epsilon_1 | c_i)
\]

\[
= \frac{1}{T} \sum_{k=t}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} r_{i,t-1-n} r_{i,s-1-l} E(e_i \epsilon_i \epsilon_m \epsilon_1 | c_i)
\]

\[
= (T-t) \frac{1}{T} \sum_{n=2}^{s-1} \sum_{l=2}^{s-1} r_{i,t-1-n} r_{i,s-1-l}.
\]

Moreover,

\[
\sum_{k=2}^{T} \sum_{m=2}^{T} E(x_{i,t-1}x_{i,s-1} \epsilon_i \epsilon_m | c_i)
\]

\[
= \sum_{k=2}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} r_{i,t-1-n} r_{i,s-1-l} E(e_i \epsilon_i \epsilon_m | c_i)
\]

\[
= \frac{1}{T} \sum_{k=2}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} r_{i,t-1-n} r_{i,s-1-l} E(e_i \epsilon_i \epsilon_m | c_i)
\]

where, since \( E(e_i \epsilon_i \epsilon_m \epsilon_1 | c_i) \) is equal to \( \sigma^4 \) for \( (k, m) = (n, l), (k, m) = (l, n) \) and \( (k, n) = (m, l) \), and zero otherwise,

\[
\sum_{k=2}^{T} \sum_{m=2}^{T} E(x_{i,t-1}x_{i,s-1} \epsilon_i \epsilon_m | c_i)
\]

\[
= \frac{1}{T} \sum_{k=2}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} r_{i,t-1-n} r_{i,s-1-l} E(e_i \epsilon_i \epsilon_m | c_i)
\]

\[
= \sigma^4 (s-2) \frac{1}{T} \sum_{k=2}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} r_{i,t-1-n} r_{i,s-1-l} + 2 \sigma^4 \frac{1}{T} \sum_{k=2}^{T} \sum_{m=2}^{T} \sum_{n=2}^{T} \sum_{l=2}^{T} r_{i,t-1-n} r_{i,s-1-l}.
\]

29
A similar calculation reveals that
\[
\sum_{k=2}^{s-1} \sum_{m=2}^{t-1} E(x_{i,t-1} x_{j,s-1} c_{ik} e_{i,m} | c_i) = \sigma^4 \frac{1}{T} \sum_{n=s}^{t-1} \sum_{j=2}^{n-1} r_{i,s-1-j} r_{i,t-1-n},
\]
and therefore
\[
\sum_{k=s}^{t-1} \sum_{m=2}^{t-1} E(x_{i,t-1} x_{j,s-1} c_{ik} e_{i,m} | c_i) = \sigma^4 (t-s-1) \frac{1}{T} \sum_{n=2}^{s-1} r_{i,t-1-n} r_{i,s-1-n},
\]
and therefore
\[
TE(x_{i,t-1} x_{j,s-1} e_i^2 | c_i) = \sigma^4 (T - 2) \frac{1}{T^2} \sum_{n=2}^{s-1} r_{i,t-1-n} r_{i,s-1-n} + 2\sigma^4 \frac{1}{T^2} \sum_{n=2}^{t-1} \sum_{l=2}^{n-1} r_{i,t-1-n} r_{i,s-1-l}
\]
\[
\to \sigma^4 \int_{u=0}^{\infty} r_i(w + v - 2u)du + 2\sigma^4 \int_{u=0}^{\infty} r_i(w + v - u - z)dudz
\]
\[
= \sigma^4 \frac{1}{2c_i \alpha} (r_i(w + v) - r_i(w - v))
\]
\[
+ 2\sigma^4 \frac{1}{(c_i \alpha)^2} (r_i(w + v) - r_i(w) - r_i(v) + 1).
\]
Thus, since
\[
\frac{1}{2c_i \alpha} \int_{w=0}^{1} \int_{v=0}^{w} (r_i(w + v) - r_i(w - v)) dv dw
\]
\[
= \frac{1}{2(c_i \alpha)^2} \left( 1 - \frac{2}{c_i \alpha} (r_i(1) - 1) + \frac{1}{2c_i \alpha} (r_i(2) - 1) \right) = \sum_{j=0}^{\infty} \frac{(2j + 1)}{2(j + 3)} \phi_j c_i^j,
\]
and
\[
\frac{2}{(c_i \alpha)^2} \int_{w=0}^{1} \int_{v=0}^{w} (r_i(w + v) - r_i(w) - r_i(v) + 1) dv dw
\]
\[
= \frac{2}{(c_i \alpha)^2} \left( \frac{1}{2} + \frac{1}{2(c_i \alpha)^2} (r_i(2) - 1) - \frac{1}{c_i \alpha} (r_i(1) - 1) - \frac{1}{(c_i \alpha)^2} (r_i(1) - 1) \right)
\]
\[
= \sum_{j=0}^{\infty} \frac{2(2j + 3 - j - 5)}{2(j + 4)(j + 3)} \phi_j c_i^j,
\]
we obtain, via symmetry,
\[
TE[(x_{i,-1} e_i^2) | c_i] = \frac{1}{T^2} \sum_{t=2}^{T} \sum_{s=2}^{T} TE(x_{i,t-1} x_{j,s-1} e_i^2 | c_i) = \frac{2}{T^2} \sum_{t=2}^{T} \sum_{s=2}^{t-1} TE(x_{i,t-1} x_{j,s-1} e_i^2 | c_i)
\]
\[
\to 2\sigma^4 \left( \frac{1}{2j + 3} \phi_j c_i^j + \frac{2(2j + 3 - j - 5)}{2(j + 4)(j + 3)} \phi_j c_i^j \right)
\]
\[
= 2\sigma^4 \sum_{j=0}^{\infty} \frac{(2j + 1)(j + 4) + 2(2j + 3 - j - 5)}{2(j + 3)(j + 4)} \phi_j c_i^j
\]
(A18)
as \(N \to \infty\).
Thus, by putting everything together,

\[
E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i-1}) \epsilon_{i,t} \right)^2 \right]
\]

\[
\rightarrow \sigma^4 \sum_{j=0}^{\infty} \phi_j \mu_j - 2\sigma^4 \sum_{j=0}^{\infty} \frac{(2j+2 + j - 1)}{2(j + 3)} \phi_j \mu_j
\]

\[
\rightarrow \sigma^4 \sum_{j=0}^{\infty} \frac{(2j+1 - 1)(j + 4) + 2(2j + 3 - j - 5)}{2(j + 3)(j + 4)} \phi_j \mu_j
\]

\[
= \sigma^4 \sum_{j=0}^{\infty} \frac{(2j+3 - 2(2j+1 - j))(j + 4) + 4(2j + 3 - j - 5)}{2(j + 3)(j + 4)} \phi_j \mu_j = \sigma^4 \omega
\]  

(A19)

as \( N \rightarrow \infty \), implying

\[
\text{var} \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i-1}) \epsilon_{i,t} \right)
\]

\[
= E \left[ \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} (x_{i,t-1} - \bar{x}_{i-1}) \epsilon_{i,t} \right)^2 \right] - \left( \frac{1}{\sqrt{T}} \sum_{t=2}^{T} E[(x_{i,t-1} - \bar{x}_{i-1}) \epsilon_{i,t}] \right)^2
\]

\[
\rightarrow \sigma^4 (\omega - \theta^2).
\]  

(A20)

Therefore, since

\[
E \left[ \left( \sqrt{T} \sum_{t=2}^{T} (r_{i,t-1} - \bar{r}_{i-1}) \epsilon_{i,t} \right)^2 \right]
\]

\[
= \tau_N E \left( (x_0)_{i_{t_0}}^2 \frac{1}{T} \sum_{t=2}^{T} (r_{i,t-1} - \bar{r}_{i-1})(r_{i,s-1} - \bar{r}_{i-1}) E(\epsilon_{i,t} \epsilon_{i,s} | c_i) \right)
\]

\[
= \sigma^2 \tau_N E \left( E[(x_0)_{i_{t_0}}^2 | c_i] \frac{1}{T} \sum_{t=2}^{T} (r_{i,t-1} - \bar{r}_{i-1})^2 \right) \rightarrow \sigma^4 \tau \gamma^*_0
\]  

(A21)

as \( N \rightarrow \infty \), we obtain

\[
\text{var} \left( \frac{1}{T} \sum_{t=2}^{T} (y_{i,t-1} - \bar{y}_{i-1}) \epsilon_{i,t} \right) \rightarrow \sigma^4 (\omega + \tau \gamma^*_0 - \theta^2),
\]  

(A22)

which in turn implies, via Theorem 2 of Phillips and Moon (1999) (the conditions of which can be verified in the same manner as in the proof of Lemma 1 (b)),

\[
A_{2N}^* - \sqrt{N} \sigma^2 \theta_N \sim \sigma^2 \sqrt{\omega + \tau \gamma^*_0 - \theta^2} N(0,1) + O_p \left( \frac{\sqrt{N}}{T} \right),
\]  

(A23)

where \( \sim \) signifies asymptotic equivalence. The proof is completed by noting that \( O_p(\sqrt{N}/T) = o_p(1) \) if \( \theta > 1/2 \) such that \( \sqrt{N}/T \rightarrow 0 \).
As for (c), by application of Corollary 1 of Phillips and Moon (1999), we have

\[ B_N^* \rightarrow_p \sigma^2 (\lambda_0^* + \tau\gamma_0^*) \tag{A24} \]

as \( N \rightarrow \infty \). This completes the proof.
Table 1: Size and power without fixed effects.

<table>
<thead>
<tr>
<th>N</th>
<th>κ = 0 t</th>
<th>κ = 0 Theory</th>
<th>κ = 0 MPP</th>
<th>κ = 1 t</th>
<th>κ = 1 Theory</th>
<th>κ = 1 MPP</th>
<th>κ = 3/2 t</th>
<th>κ = 3/2 Theory</th>
<th>κ = 3/2 MPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>5.3</td>
<td>4.5</td>
<td>4.5</td>
<td>5.2</td>
<td>4.6</td>
<td>4.6</td>
<td>4.7</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>80</td>
<td>5.4</td>
<td>4.6</td>
<td>4.6</td>
<td>5.1</td>
<td>4.5</td>
<td>4.5</td>
<td>4.8</td>
<td>4.9</td>
<td>4.9</td>
</tr>
<tr>
<td>160</td>
<td>5.4</td>
<td>5.3</td>
<td>5.3</td>
<td>5.9</td>
<td>4.9</td>
<td>4.9</td>
<td>5.7</td>
<td>5.1</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>a = b = 0</td>
<td></td>
<td></td>
<td>a = b = −2</td>
<td>25.5</td>
<td>24.2</td>
<td>28.8</td>
<td>52.5</td>
<td>51.2</td>
</tr>
<tr>
<td>40</td>
<td>26.5</td>
<td>25.5</td>
<td>28.8</td>
<td>57.7</td>
<td>55.1</td>
<td>29.5</td>
<td>86.9</td>
<td>85.6</td>
<td>30.0</td>
</tr>
<tr>
<td>80</td>
<td>28.7</td>
<td>27.7</td>
<td>30.3</td>
<td>60.7</td>
<td>57.5</td>
<td>29.7</td>
<td>95.2</td>
<td>94.1</td>
<td>29.0</td>
</tr>
<tr>
<td></td>
<td>a = b = −4</td>
<td></td>
<td></td>
<td>a = b = −6, b = −2</td>
<td>71.9</td>
<td>64.2</td>
<td>81.4</td>
<td>94.7</td>
<td>90.8</td>
</tr>
<tr>
<td>40</td>
<td>75.5</td>
<td>69.2</td>
<td>80.4</td>
<td>97.4</td>
<td>95.6</td>
<td>81.9</td>
<td>99.4</td>
<td>98.9</td>
<td>82.1</td>
</tr>
<tr>
<td>80</td>
<td>76.5</td>
<td>72.5</td>
<td>79.7</td>
<td>98.7</td>
<td>97.7</td>
<td>80.2</td>
<td>100.0</td>
<td>99.9</td>
<td>80.1</td>
</tr>
<tr>
<td>160</td>
<td>69.2</td>
<td>62.0</td>
<td>81.4</td>
<td>92.0</td>
<td>88.0</td>
<td>80.9</td>
<td>95.6</td>
<td>92.3</td>
<td>79.9</td>
</tr>
<tr>
<td></td>
<td>a = −6, b = −2</td>
<td></td>
<td></td>
<td>a = −8, b = 0</td>
<td>72.6</td>
<td>67.4</td>
<td>80.4</td>
<td>96.4</td>
<td>94.2</td>
</tr>
<tr>
<td>40</td>
<td>75.0</td>
<td>71.3</td>
<td>79.7</td>
<td>98.3</td>
<td>97.1</td>
<td>80.2</td>
<td>99.9</td>
<td>99.8</td>
<td>80.1</td>
</tr>
<tr>
<td>80</td>
<td>70.6</td>
<td>67.2</td>
<td>79.7</td>
<td>94.6</td>
<td>93.3</td>
<td>80.2</td>
<td>96.0</td>
<td>95.2</td>
<td>80.1</td>
</tr>
</tbody>
</table>

Notes: The data is generated as $y_{it} = \rho_i y_{i,t-1} + \epsilon_{it}$, where $t = -T_0 + 2, ..., T, i = 1, ..., N, y_{i,T_0+1} = 0, \epsilon_{it} \sim N(0, 1), \rho_i = \exp(c_i/N^\eta T), c_i \sim U(a, b), T = N$ and $T_0 = \lceil T^\kappa \rceil$. $t_1$ refers to the actual $t$-test, “Theory” refers to the asymptotic power function derived in Section 3 and “MPP” refers to the first-order power function that assumes $\kappa = 0$, as given in Moon et al. (2007). All tests are conducted at the 5% level.
Table 2: Size and power with fixed effects.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$t^*$</th>
<th>Theory</th>
<th>MMP</th>
<th>$t_0^*$</th>
<th>Theory</th>
<th>$t_2^*$</th>
<th>Theory</th>
<th>MMP</th>
<th>$t_0^*$</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>6.8</td>
<td>4.5</td>
<td>4.5</td>
<td>6.8</td>
<td>4.5</td>
<td>5.7</td>
<td>4.6</td>
<td>4.6</td>
<td>5.7</td>
<td>4.6</td>
</tr>
<tr>
<td>80</td>
<td>5.7</td>
<td>4.6</td>
<td>4.6</td>
<td>5.7</td>
<td>4.6</td>
<td>5.1</td>
<td>4.7</td>
<td>4.7</td>
<td>5.1</td>
<td>4.7</td>
</tr>
<tr>
<td>160</td>
<td>4.9</td>
<td>5.3</td>
<td>5.3</td>
<td>4.9</td>
<td>5.3</td>
<td>4.6</td>
<td>4.7</td>
<td>4.7</td>
<td>4.6</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Notes: $t^*$ and $t_0^*$ refer to the feasible and infeasible $t$-tests, respectively. See Table 1 for an explanation of the rest.