On the Importance of the First Observation in GLS Detrending in Unit Root Testing

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 Please do not quote without obtaining the author’s consent as these works are in their draft form.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.
ON THE IMPORTANCE OF THE FIRST OBSERVATION IN GLS DETRENDING IN UNIT ROOT TESTING

June 6, 2013

Abstract

First-differencing is generally taken to imply the loss of one observation, the first, or at least that the effect of ignoring this observation is asymptotically negligible. However, this is not always true, as in the case of GLS detrending. In order to illustrate this, the current paper considers as an example the use of GLS detrended data when testing for a unit root. The results show that the treatment of the first observation is absolutely crucial for test performance, and that ignorance causes test break-down.

JEL Classification: C12; C13; C33.

Keywords: Unit root test; GLS detrending; Local asymptotic power.

1 Introduction

Consider the time series variable \( \{X_s\}_{s=1}^t \). The present paper originates with the following very basic question: how should we define \( \Delta X_s \), where \( \Delta X_s = X_s - X_{s-1} \)? In particular, while the definitions of \( \Delta X_2, ..., \Delta X_t \) are clear, it is less obvious how to treat \( \Delta X_1 \). One way is to say that, since accumulation should undo first-differencing, we must have \( \Delta X_1 = X_1 \), because only then will it be true that \( \sum_{s=1}^t \Delta X_s = X_t \). Hence, according to this, we have

\[
\Delta X_s = X_s - X_{s-1} \quad \text{for} \quad s = 2, ..., t \quad \text{and} \quad \Delta X_1 = X_1.
\]

However, this definition is probably more the exception rather than the rule. Indeed, the by far most common approach, is to simply ignore \( \Delta X_1 \), as when defining

\[
\Delta X_s = X_s - X_{s-1} \quad \text{for} \quad s = 2, ..., t \quad \text{and} \quad \Delta X_1 = 0,
\]

in which case \( \sum_{s=1}^t \Delta X_s = \sum_{s=2}^t \Delta X_s = X_t - X_1 \). The obvious rationale for this in the unit root case is that since the relevant quantity here is not \( \sum_{s=1}^t \Delta X_s \) but rather \( T^{-1/2} \) times this
sum, the effect of ignoring \( \Delta X_1 \) is negligible; \( T^{-1/2} \sum_{s=2}^{T} \Delta X_s = T^{-1/2} (X_t - X_1) = T^{-1/2} X_t + o_p(1) \). This is logical; if the data are \( O_p(\sqrt{T}) \), as in the unit root case, the treatment of the first observation, which is only \( O_p(1) \), should not matter.

One situation when this issue becomes relevant is when performing “generalized least squares (GLS) detrending” (Elliott et al., 1996). In this case, the relevant (quasi) differenced variable is given by \( \Delta \varphi X_t = X_t - \varphi X_{t-1} \), where \( \varphi \) is such that \( \varphi = 1 + T^{-1/2} \). The appropriate value of \( \varphi \) to consider depends on the deterministic specification, and is given in Elliott et al. (1996). The obvious question here is: how to treat \( \Delta \varphi X_1 \)? Elliott et al. (1996) adopt definition (1) and set \( \Delta \varphi X_1 = X_1 \). The resulting GLS detrended variable is given by \( X^\text{gls}_t = X_t - \hat{\beta}' D_t \), where \( D_t \) is a vector of deterministic constant and trend terms, and \( \hat{\beta} \) is obtained from an ordinary least squares (OLS) regression of \( \Delta \varphi X_t \) onto \( \Delta \varphi D_t \).

The idea now is to simply use \( X^\text{gls}_t \) as input data when testing for a unit root. Elliott et al. (1996) considers the Dickey–Fuller (DF) test, which they refer to as “DF–GLS” to stress the fact that the data have been detrended using GLS rather than OLS (as originally proposed). The motivation for making this change is that since in GLS the detrending is performed under a local/near unit root alternative, the resulting test should have relatively high power against such alternatives, a result that has been confirmed by numerous Monte Carlo studies. The use of GLS detrended data has since then become very popular. In fact, it is almost difficult to find a recent contribution in the unit root literature that is not based on GLS (see Ng and Perron, 2001; Elliott and Jansson, 2003; Müller and Elliott, 2003; Carrion-i-Silvestre et al., 2009; Harvey et al., 2009, to mention a few).

Another possibility that is expected to lead to even higher power is to combine GLS detrending with panel data. Interestingly, while the power increasing potential of such a combination has been widely recognized (see, for example, Breitung and Pesaran, 2008; Breuer et al., 2002; Bai and Ng, 2004, 2010; Choi, 2001; Lopez, 2009; Moon and Perron, 2008), the only real in-depth study known to us is that of Lopez (2009), who uses Monte Carlo simulation to study the small-sample properties of her pooled GLS-based \( t \)-test for a unit root. The results suggest that when definition (1) is used the resulting test is well-behaved with good size accuracy and higher power than the corresponding OLS-based test, thus corroborating the results in the time series case.

The purpose of this study is to investigate the effect of a change in the definition of \( \Delta \varphi X_t \), from (1) to (2). In view of the above discussion, it seems reasonable to assume that the effect
is negligible, and hence that one can just as well use (2) and simply ignore $\Delta_0 X_1$ (and $\Delta_0 D_1$). The most natural candidate for a test statistic is DF–GLS, and in this paper we therefore start by considering this test statistic. However, we also consider panel data. The usual way in which GLS-based tests are implemented is to take any existing OLS-based test statistic, and to just apply it to the GLS demeaned data (as in the case of DF–GLS), here under (2). We do the same. The two most common panel unit root test approaches for use in the baseline case with fixed effects, denoted $t^#$ and $t^+$, are both based on the pooled OLS $t$-statistic for a unit root, but differ in the way that the test statistic is corrected for the so-called “Nickell bias”, which is an artifact of the demeaning. In particular, while $t^#$ is based on correcting only the numerator, in $t^+$ the entire test statistic is corrected (see Moon and Perron, 2008). We consider both.

Given the good performance of the GLS-based test statistics under (1), and the fact that the treatment of the first observation should not matter, the test statistics under (2) considered here are expected to behave very similarly. However, this is not what we find. In fact, all three test statistics diverge with the sample size, and this is true under both the null and local alternative hypotheses. The obvious implication for practical work involving GLS detrended data is to always use (1).

2 Model

Consider the panel data variable $X_{i,t}$, observable for $t = 1, \ldots, T$ time series and $i = 1, \ldots, N$ cross-sectional units. The data generating process (DGP) of this variables is assumed to be given by

\begin{align*}
X_{i,t} & = \theta_i + U_{i,t}, \\
U_{i,t} & = \rho_i U_{i,t-1} + \epsilon_{i,t},
\end{align*}

where $U_{i,0} = 0$, $\epsilon_{i,t}$ is independent and identically distributed (iid) with $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,t}^2) = \sigma_i^2 > 0$ and $E(\epsilon_{i,t}^4) < \infty$, and $\theta_i$ is a unit-specific fixed effect. It is further assumed that

$$
\rho_i = \exp(N^{-k} T^{-1} c_i),
$$

\footnote{In Moon and Perron (2008), $t^#$ and $t^+$ refer to specific test statistics, which is also how we refer to them here. However, they can also be viewed more broadly as representing the two classes of tests that arise as a result of the difference in bias correction. The “$t^+$ class” is broadest and includes the $t^\ast_d$ statistic of Levin et al. (2002), and the Bai and Ng (2010) $P_a$ and $P_b$ statistics. The $t^\ast_a$ and $t^\ast_b$ statistics of Moon and Perron (2004) belong to the “$t^#$ class”.}
where $\kappa \geq 0$ and $c_i$ is iid with at least two finite moments, which are henceforth going to be denoted as $\mu_j = E(c_i^j)$ for $j \geq 1$ and $\mu_0 = 0$. Also, $c_i$ and $\epsilon_{i,t}$ are mutually independent. The appropriate formulation of the null and alternative hypotheses depends on the test statistic being considered. In case of DF–GLS when applied to unit $i$, henceforth denoted $t_{gls,i}$, $H_0 : c_i = 0$ is tested versus $H_1 : c_i < 0$, while in case of the GLS-versions of $t^\#$ and $t^+$, henceforth denoted $t^\#_{gls}$ and $t^+_{gls}$, respectively, $H_0 : \mu_2 = 0$ (or $c_1 = \ldots = c_N = 0$) is tested versus $H_1 : \mu_2 > 0$ (or $c_i \neq 0$ for at least some $i$).

This completes our DGP. We would like end this section by pointing out that while admittedly very simple, in the present study there seem to be little or no merit in considering a more general DGP. Indeed, the introduction of nuisance parameters is hardly going to make the GLS-based tests considered here behave any better. Hence, if the tests do not perform here, the chances of them performing better in a more general DGP are very small, if not nonexistent. In fact, as is well-known from the time series literature (see Müller and Elliott, 2003), the relatively good performance of GLS under (1) depends on, among other things, the extent of the initialization, and that it is likely to deteriorate with increases in $|U_{i,0}|$. Hence, if anything, the consideration of a more general DGP is expected to lead to even poorer performance.

3 Results

The detrending (under (2)) is carried out as described in Section 1 with $D_t = 1$ (to account for the fixed effect) and

$$\bar{p}_t = 1 + N^{-\kappa T^{-1}} \bar{\eta}_t$$  \hspace{1cm} (6)

in place of $\bar{p}$.\(^2\) While in the time series case the GLS demeaning is typically carried out while setting $\bar{\kappa} = 0$, in panels the choice of $\bar{\kappa}$ is less obvious. For example, while Lopez (2009) sets $\bar{\kappa} = \kappa$, Choi (2001) sets $\bar{\kappa} = 0$. Because of this, in the present study the appropriate value of $\bar{\kappa}$ is considered a part of the analysis. The drift parameter $\bar{\eta}_t$ can be either random or non-random. In both cases, the main assumption is that $\prod_{\kappa} < \infty$, where $\bar{\eta}_j = N^{-1} \sum_{i=1}^N E(\bar{\eta}_i^j)$. Also, for simplicity, and since we are not specifically interested in the regularity conditions under which consistency holds, in this paper we simply assume that a consistent estimator

\(^2\)Note how $\rho_i = \exp(N^{-\kappa T^{-1}} c_i) = 1 + N^{-\kappa T^{-1}} c_i + o_p(1)$. The specification in (6) therefore mimics that in (5).
$\hat{\sigma}^2$ of $\sigma^2$ is available. For now we therefore assume $(\hat{\sigma}^2 - \sigma^2) = o_p(1)$; later on we describe how $\hat{\sigma}^2$ may be constructed in practice.\footnote{This is less demanding than Assumption 3.1 of Moon and Perron (2008), requiring that $(\hat{\sigma}^2 - \sigma^2) = o_p(N^{-1/2})$.}

3.1 The time series statistic

Denote by $\hat{\alpha}_{\text{gls},i}$ the OLS slope estimator in a time series regression of $\Delta X_{gls,i}^{t}$ onto $X_{gls,i}^{t-1}$. The $t_{\text{gls},i}$ statistic is given by

$$t_{\text{gls},i} = \frac{\hat{\alpha}_{\text{gls},i}}{\sqrt{\sum_{t=2}^{T} (X_{\text{gls},i}^{t-1})^2}},$$

whose asymptotic distribution is given in Theorem 1.

**Theorem 1.** Under the conditions stated above, as $T \to \infty$,

$$N^{-\frac{3}{2}}T^{-1/2}t_{\text{gls},i} \to_d \frac{\omega_i \sqrt{2}N(0,1)}{\sqrt{\omega_i^2 \chi^2(1) + \omega_i^4 N^{-2} \int_{r=0}^{1} |f_i^{\text{ols}}(r)|^2 dr}},$$

where $\to_d$ signifies convergence in distribution, $\omega_i^2 = \int_{r=0}^{1} \exp[2(1-r)c_i N^{-\kappa}]dr$, $f_i^{\text{ols}}(r) = J_i(r) - \int_{s=0}^{r} J_i(s)ds$, $J_i(r) = \int_{s=0}^{r} \exp[(r-s)c_i N^{-\kappa}]dW_i(s)$, $W_i(r)$ is a standard Brownian motion, and $\chi^2(1)$ is a chi-squared variate with one degree of freedom.

As Elliott et al. (1996) show, the asymptotic distribution of $t_{\text{gls},i}$ under (1) is given by

$$c_i \sqrt{\int_{r=0}^{1} |J_i(r)|^2 dr} + \int_{r=0}^{1} J_i(r)dW_i(r) \sqrt{\int_{r=0}^{1} |J_i(r)|^2 dr},$$

(7)

which is identically the local asymptotic distribution of the conventional DF test statistic in the case without deterministic constant and trend terms. Hence, under (1) the test statistic has a well-defined limiting distribution. By contrast, since according Theorem 1, $N^{-\frac{3}{2}}T^{-1/2}t_{\text{gls},i} = O_p(1)$, we have $t_{\text{gls},i} = O_p(N^{-\frac{3}{2}} \sqrt{T})$. Hence, while well-behaved under (1), if the GLS demeaning is done under (2), this makes the test statistic divergent. In the proof of Theorem 1 is given in a supplement to this paper. In Remark 2 of that supplement we show that the divergence is due to the fact that $\sqrt{T}\hat{\alpha}_{\text{gls},i} = O_p(1)$. Hence, quite surprisingly the use of GLS rather than OLS demeaning actually leads to a reduction in the rate of consistency of $\hat{\alpha}_{\text{gls},i}$ from the usual unit root (or “superconsistency”) rate of $T$ to $\sqrt{T}$ (which is the usual rate in the case of stationary data). Thus, contrary to intuition, the first observation is
not negligible, but is in fact absolutely crucial for test performance. Note also that while one could in principle chose to ignore the first observation and use $\tilde{t}_{gls,i} = N^{-\frac{1}{2}}t_{gls,i}$ as a test statistic, this is not recommended. The main reason is that, in contrast to (7), the asymptotic distribution under (1), in which $c_i$ entails a shift in both the mean and variance, according to Theorem 2, under (2) $c_i$ has only a variance effect, which is suggestive of low power, as is to be expected given the relatively low rate of consistency of $\hat{a}_{gls,i}$ in this case.

**Remark 1.** The denominator of the asymptotic distribution of $N^{-\frac{1}{2}}t_{gls,i}$ can be seen as a weighted sum of a chi-squared variate and an integral that is identically the denominator in the asymptotic distribution of the usual ADF test statistic based on OLS demeaning. Both depend on $c_i$ and on the rate of shrinking of the local alternative, as measured by $k$. If $k > 0$, then $\omega_i^2 = 1 + o_p(1)$ and $J_i(r) = W_i(r) + o_p(1)$ as $N \to \infty$, and therefore the dependence of these quantities on $c_i$ disappears. In an essence, if $k > 0$, we are too close to the null for the test to have power. If, on the other hand, $k = 0$, then $\omega_i^2 = \int_0^r \exp[(r-s)c_i]dW_i(s)$, and so power is non-negligible.

**Remark 2.** The rate of shrinking of $\tilde{p}_i$, as captured by $\tilde{\kappa}$, is important for the shape of the asymptotic distribution of $N^{-\frac{1}{2}}t_{gls,i}$. Note in particular that if $\tilde{\kappa} > 0$, as $N \to \infty$, the asymptotic distribution of $N^{-\frac{1}{2}}t_{gls,i}$ reduces to $\sqrt{2}N(0,1)/\sqrt{\chi^2(1)}$, which is independent of $c_i$, and therefore power is negligible.

### 3.2 The panel statistics

Denote by $\hat{a}_{gls}$ the pooled OLS slope estimator in a panel regression of $\Delta X_{gls,i,t}$ onto $X_{gls,i,t-1}$. The panel test statistics that we consider are the following GLS demeaned versions of the (OLS demeaned) $t^#$ and $t^+$ statistics of Moon and Perron (2008):

$$t^#_{gls} = \frac{\hat{a}_{gls}^#}{\hat{\sigma}_x / \sqrt{\sum_{i=1}^N \sum_{t=2}^T (X_{gls,i,t-1})^2}}$$

$$t^+_{gls} = \frac{\hat{a}_{gls}^+}{\hat{\sigma}_x / \sqrt{\sum_{i=1}^N \sum_{t=2}^T (X_{gls,i,t-1})^2}}$$

where

$$\hat{a}_{gls}^# = \hat{a}_{gls} + \frac{NT\hat{\sigma}_x^2}{2 \sum_{i=1}^N \sum_{t=2}^T (X_{gls,i,t-1})^2}$$

$$\hat{a}_{gls}^+ = \hat{a}_{gls} + \frac{3}{T}$$
As is well-known, demeaning induces a correlation between the disturbance and the lagged dependent variable, rendering the resulting OLS estimator of the autoregressive coefficient biased. This is the Nickell bias. In the classical “micro” panel setting with $T$ fixed and $N \to \infty$ this bias is rather devastating, as in this case OLS is even inconsistent. The problem is made less severe by assuming that $N, T \to \infty$; however, while consistent, the asymptotic distribution of the OLS estimator is still miscentered, which in turn calls for some kind of correction. This is why $t^\#_\text{gls}$ and $t^+_\text{gls}$ are based on $\hat{\alpha}^\#_\text{gls}$ and $\hat{\alpha}^+_\text{gls}$, respectively, and not on $\hat{\alpha}_\text{gls}$. The difference between the two is that while $\hat{\alpha}^\#_\text{gls}$ only corrects the numerator, in $\hat{\alpha}^+_\text{gls}$ the entire test statistic is corrected.

We start by considering $t^\#_\text{gls}$, the local asymptotic distribution of which is provided in Theorem 2.

**Theorem 2.** Under the conditions stated above, as $N, T \to \infty$ with $\sqrt{N}/T = o(1)$ and $N/T = O(1)$,

$$N^{-\kappa}T^{-1/2}t^\#_\text{gls} \sim -\frac{\sqrt{N}T^{-1/2}\mu_2}{\sqrt{\mu_2 + N^{-2\kappa}/6}} + \left(\frac{2\mu_4}{\mu_2 + N^{-2\kappa}/6}\right)^{1/2} N(0,1),$$

where $\sim$ signifies asymptotic equivalence.

As Theorem 2 makes clear, the asymptotic distribution of $N^{-\kappa}T^{-1/2}t^\#_\text{gls}$ can be divided into two terms; (i) a constant drift term that captures the dependence on $\mu_2$ and that is there also under the unit root null, and (ii) a $N(0,1)$ variable. Since by assumption $N/T = O(1)$, both terms are $O_p(1)$, and therefore $t^\#_\text{gls} = O_p(N^{\kappa}\sqrt{T})$. Hence, analogous to the results reported in Section 3.1 for $t^\text{gls}$, $t^\#_\text{gls}$ is divergent, which is of course very much unlike the case of OLS demeaning. Note also how the asymptotic distribution of $N^{-\kappa}T^{-1/2}t^\#_\text{gls}$ is independent of $c_i$ and its moments. Hence, even if appropriately normalized to have a nuisance parameter free distribution under the null, the power of resulting test statistic,

$$t^\#_\text{gls} = \left(\frac{\mu_2 + N^{-2\kappa}/6}{2\mu_4}\right)^{1/2} N^{-\kappa}T^{-1/2}t^\#_\text{gls} + \frac{\sqrt{N}T^{-1/2}\mu_2}{\sqrt{2\mu_4}},$$

would be negligible, and this is true for all $\kappa > 0$. This is again very different from the case of OLS demeaned data, in which $t^\#$ has non-negligible (and non-increasing) local power for $\kappa = 1/4$ (see Moon and Perron, 2008, Theorem 4.1). Hence, as in the time series case, while one could in principle consider ignoring the first observation, and to use $t^\#_\text{gls}$ as a test statistic, this is not recommended.
Remark 3. By using the results provided in the proof of Theorem 1 (see the supplemental material), it is possible to show that
\[ \sqrt{NT} \hat{a}_gls^# = \sqrt{NT} \hat{a}_{gls} + O_p(N^{1/2-2\tau}T^{-1/2}). \]
Hence, provided that \( N^{1/2-2\tau}T^{-1/2} = o(1) \), \( \hat{a}_gls^# \) and \( \hat{a}_{gls} \) not only have the same asymptotic distribution but are in fact asymptotically equivalent. The implication is that, again unlike the case with OLS demeaning, as long as \( N^{1/2-2\tau}T^{-1/2} = o(1) \), with GLS demeaning there actually no need for any bias correction.

Remark 4. In the proof of Theorem 2 (see the supplemental material), we show that
\[ \hat{t}_{gls}^# \sim \frac{N^{1/2-2(\kappa+\tau)}T^{-1/2}\mu_2}{24\sqrt{2\mu^4_1}} + N(0,1) + O_p(cN^{-k}), \]
where \( O_p(cN^{-k}) \) is a first-order term in \( c_iN^{-k} \). Note how the first term on the right-hand side depends on \( \mu_2 \), which drives power. Hence, while negligible for \( \kappa > 0 \), if we assume that \( \kappa = 0 \) and \( N/T \to \tau > 0 \), such that \( N^{1/2-2(\kappa+\tau)}T^{-1/2} \to \tau \), local power is non-negligible. In other words, for this test to have power it has to be set up against \( \rho_i = \exp(c_iT^{-1}) \), which is nothing but the conventional time series local alternative. We therefore come to the somewhat counterintuitive conclusion that in terms of local power, the use of the information contained in the cross-sectional dimension does not add anything. That is, one can just as well use a time series test (which would have power in the same neighborhood around unity).

Let us now consider the asymptotic distribution of \( t_{gls}^+ \), which is given in Theorem 3.

Theorem 3. Under the conditions of Theorem 1,
\[ N^{-\kappa}T^{-1/2}t_{gls}^+ = N^{-\kappa}T^{-1/2}t_{gls}^# + \frac{3\sqrt{N^{1/2}\mu_2}}{\sqrt{\mu_2} + N^{2\kappa}/6} + o_p(1). \]

Moon and Perron (2008, Theorems 4.1 and 4.2) show that if \( \sqrt{N}/T = o(1) \), while \( t^+ \) has non-negligible local power for \( \kappa = 1/2 \), \( t^# \) does not. In fact, for \( t^# \) to have power one has to set \( \kappa = 1/4 \), which is suggestive of relatively low power (see Moon and Perron, 2008, page 81). Moreover, while the power of \( t^# \) is driven by \( \mu_2 \) (which includes both the mean and variance of \( c_i \)), the power of \( t^+ \) is driven only by \( \mu_1 \). The power properties of these tests are therefore very different. Contrasting this, Theorem 3 shows that if \( N/T = o(1) \),
$N^{-\kappa} T^{-1/2} t_{gls}^+$ and $N^{-\kappa} T^{-1/2} t_{gls}^#$ are asymptotically equivalent tests. Thus, since the local power of $N^{-\kappa} T^{-1/2} t_{gls}^#$ is negligible (Theorem 2), the power of $N^{-\kappa} T^{-1/2} t_{gls}^+$ is negligible too. In fact, since $3\sqrt{N T^{-1/2} \bar{\mu}_2 / \sqrt{\mu}_2 + N^{-2\kappa} / 6}$ (the difference between the two tests) is independent of $c_i$, the local power of the appropriately centered and scaled version of $N^{-\kappa} T^{-1/2} t_{gls}^+$, $\tilde{t}_{gls}^+$ say, is negligible even if $N/T = O(1)$ does not go to zero.

Remark 5. In the proof of Theorem 3 (see the supplement), we show that if $\kappa = 0$ is permitted, then

$$N^{-\kappa} T^{-1/2} t_{gls}^+ = N^{-\kappa} T^{-1/2} t_{gls}^# + \frac{N^{1/2 - (\kappa + 2\kappa) T^{-1/2} \mu_1}}{4 \sqrt{\mu}_2 + N^{-2\kappa} / 6} + \frac{3\sqrt{N T^{-1/2} \bar{\mu}_2}}{\sqrt{\mu}_2 + N^{-2\kappa} / 6} + O_p(cN^{-\kappa}).$$

Thus, just as before when looking at $N^{-\kappa} T^{-1/2} t_{gls}^#$ (see Remark 4), relaxing of the assumption that $\kappa > 0$ leads to a change in the results, as captured by the second term on the right-hand side, which is activated when $\bar{\kappa} = \kappa = 0$ and $N/T \rightarrow \tau > 0$ as $N, T \rightarrow \infty$. However, unlike before, although there is a dependence on $\mu_1$, in this instance there is no dependence on higher moments, such as $\mu_2$. Another implication is that while the drift term in $N^{-\kappa} T^{-1/2} t_{gls}^#$ is $O(N^{1/2 - (\kappa + \kappa) T^{-1/2}})$ (see again Remark 4), the drift in $N^{-\kappa} T^{-1/2} t_{gls}^+$ is $O(N^{1/2 - (\kappa + 2\kappa) T^{-1/2}}) > O(N^{1/2 - (\kappa + \kappa) T^{-1/2}})$. This means that in small samples the power of $N^{-\kappa} T^{-1/2} t_{gls}^+$ is expected to be larger than that of $N^{-\kappa} T^{-1/2} t_{gls}^#$, which is in line with the results for the OLS demeaned versions of these tests (see the discussion following Theorem 3).

3.3 Monte Carlo simulations

In this section we report the results of a small-scale simulation study. The purpose is to assess the small-sample accuracy of our theoretical results only; none of the tests considered here should be used in practice. Because of this, for simplicity, we focus on the panel statistics; some results for the time series statistic are available upon request.

The DGP is given by a restricted version of (3)–(5) that sets $\epsilon_{it} \sim N(0, 1), \theta_i = 1, \bar{c}_i = \bar{c} = -1$ and $c_i = c$. The results for $\tilde{r}_{gls}^#$ and $\tilde{t}_{gls}^+$ are compared with those obtained by applying the OLS demeaned $t^#$ and $t^+$ statistics of Moon and Perron (2008). All statistics are implemented with $\hat{\sigma}_e^2$ set equal to $\hat{\sigma}_e^2 = (NT)^{-1} \sum_{t=1}^{N} \sum_{t=2}^{T} (\Delta x_{it})^2$. The significance level is set to 5% and all tests are double-sided. The number of replications is set to 3,000.

The results reported in Table 1 are generally in agreement with theory and can be summarized as follows. First, the size accuracy of the tests is generally good. There are some
distortions but these diminish as \( N \) and \( T \) increases. Second, the power of the GLS-based tests is generally close to size. One exception is when \( \rho = \kappa = 0 \) and \( N = T \), which is just as expected given Remark 4. We also see that while \( \bar{t}_{gls}^+ \) tend to have some power also when \( \kappa > 0 \), this is mainly a small-sample effect that goes away when \( N \) and \( T \) increases. Thus, while for the sample sizes considered here \( \bar{t}_{gls}^+ \) tend to be more powerful than \( t^\# \), as \( N \) and \( T \) increases the relative power of \( t^\# \) will tend to increase, eventually becoming more powerful. Table 1 is quite suggestive of this. Third, the fact when \( \kappa > 0 \) \( \bar{t}_{gls}^+ \) is generally more powerful than \( \bar{t}_{gls}^\# \) is in accordance with our theoretical prediction (see Remark 5). The overall best power is obtained by using \( t^+ \) (as expected; see the discussion following Theorem 3). Fourth, while the power of \( t^+ (t^\#) \) is roughly constant in \( N \) and \( T \) when \( \kappa = 1/2 \) (\( \kappa = 1/4 \)), as expected, for \( \kappa < 1/2 \) (\( \kappa = 0 \)) power is increasing in \( N \) and \( T \). We similarly see that while the powers of \( \bar{t}_{gls}^\# \) and \( \bar{t}_{gls}^+ \) do not vary much when \( \kappa = 0 \), when \( \kappa > 0 \) it is decreasing in the sample size.

4 Conclusion

The present paper shows how the treatment of the first observation in GLS demeaning can have a huge impact on test performance when testing for a unit root. As examples, we consider both the conventional DF–GLS test of Elliott et al. (1996), and GLS versions of the panel data unit root tests previously considered by Moon and Perron (2008) in the case of OLS demeaning. Our results show that if the GLS demeaning is carried out while ignoring the first observation (as under (2)), then all three tests break down and actually become divergent, and this is true under both the null and alternative hypotheses. The implication is that the usual argument that the effect of the first observation is negligible does not apply, and that practitioners need to be aware of this.
References


Table 1: Size and power at the 5% level.

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$t_{\text{gls}}^*$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>4.6</td>
<td>4.9</td>
<td>4.7</td>
<td>4.9</td>
<td>3.0</td>
<td>5.4</td>
<td>1.0</td>
</tr>
<tr>
<td>15</td>
<td>225</td>
<td>5.5</td>
<td>5.6</td>
<td>5.1</td>
<td>3.8</td>
<td>2.8</td>
<td>4.7</td>
<td>1.3</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>4.1</td>
<td>4.3</td>
<td>4.6</td>
<td>4.4</td>
<td>3.1</td>
<td>4.2</td>
<td>1.1</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>8.2</td>
<td>8.9</td>
<td>8.6</td>
<td>3.5</td>
<td>6.8</td>
<td>30.8</td>
<td>4.1</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>6.5</td>
<td>6.8</td>
<td>5.6</td>
<td>4.9</td>
<td>4.4</td>
<td>31.6</td>
<td>6.0</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>5.8</td>
<td>6.0</td>
<td>5.7</td>
<td>4.9</td>
<td>3.8</td>
<td>30.8</td>
<td>16.8</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>4.8</td>
<td>4.7</td>
<td>5.4</td>
<td>4.7</td>
<td>3.3</td>
<td>33.6</td>
<td>46.8</td>
</tr>
</tbody>
</table>

$c = \kappa = \bar{\kappa} = 0$

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$t_{\text{gls}}^*$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>4.4</td>
<td>5.3</td>
<td>1.5</td>
<td>15.2</td>
<td>4.6</td>
<td>5.4</td>
<td>2.5</td>
</tr>
<tr>
<td>15</td>
<td>225</td>
<td>4.6</td>
<td>5.5</td>
<td>1.8</td>
<td>20.0</td>
<td>5.0</td>
<td>5.4</td>
<td>2.7</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>4.3</td>
<td>4.8</td>
<td>1.7</td>
<td>23.9</td>
<td>4.4</td>
<td>4.6</td>
<td>2.6</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>8.7</td>
<td>17.0</td>
<td>5.0</td>
<td>20.2</td>
<td>8.8</td>
<td>11.8</td>
<td>6.0</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>6.6</td>
<td>13.3</td>
<td>3.7</td>
<td>41.9</td>
<td>6.4</td>
<td>8.0</td>
<td>4.5</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>5.8</td>
<td>10.2</td>
<td>4.4</td>
<td>59.7</td>
<td>5.8</td>
<td>6.5</td>
<td>4.3</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>4.8</td>
<td>9.5</td>
<td>5.2</td>
<td>80.6</td>
<td>4.6</td>
<td>5.2</td>
<td>5.2</td>
</tr>
</tbody>
</table>

$c = -2, \kappa = \bar{\kappa} = 1/4$

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$t_{\text{gls}}^*$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>4.4</td>
<td>5.3</td>
<td>1.5</td>
<td>15.2</td>
<td>4.6</td>
<td>5.4</td>
<td>2.5</td>
</tr>
<tr>
<td>15</td>
<td>225</td>
<td>4.6</td>
<td>5.5</td>
<td>1.8</td>
<td>20.0</td>
<td>5.0</td>
<td>5.4</td>
<td>2.7</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>4.3</td>
<td>4.8</td>
<td>1.7</td>
<td>23.9</td>
<td>4.4</td>
<td>4.6</td>
<td>2.6</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>8.7</td>
<td>17.0</td>
<td>5.0</td>
<td>20.2</td>
<td>8.8</td>
<td>11.8</td>
<td>6.0</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>6.6</td>
<td>13.3</td>
<td>3.7</td>
<td>41.9</td>
<td>6.4</td>
<td>8.0</td>
<td>4.5</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>5.8</td>
<td>10.2</td>
<td>4.4</td>
<td>59.7</td>
<td>5.8</td>
<td>6.5</td>
<td>4.3</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>4.8</td>
<td>9.5</td>
<td>5.2</td>
<td>80.6</td>
<td>4.6</td>
<td>5.2</td>
<td>5.2</td>
</tr>
</tbody>
</table>

$c = -4, \kappa = \bar{\kappa} = 0$

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>$t_{\text{gls}}^*$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
<th>$t_{\text{gls}}^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>100</td>
<td>2.6</td>
<td>6.9</td>
<td>0.3</td>
<td>75.2</td>
<td>4.1</td>
<td>6.1</td>
<td>1.0</td>
</tr>
<tr>
<td>15</td>
<td>225</td>
<td>1.7</td>
<td>4.7</td>
<td>0.4</td>
<td>92.7</td>
<td>4.2</td>
<td>5.5</td>
<td>1.3</td>
</tr>
<tr>
<td>20</td>
<td>400</td>
<td>2.1</td>
<td>4.8</td>
<td>0.5</td>
<td>98.2</td>
<td>4.0</td>
<td>5.3</td>
<td>1.1</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>6.4</td>
<td>51.4</td>
<td>1.6</td>
<td>94.4</td>
<td>8.3</td>
<td>25.2</td>
<td>4.1</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>3.4</td>
<td>54.5</td>
<td>3.7</td>
<td>100.0</td>
<td>6.5</td>
<td>21.4</td>
<td>4.8</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>3.2</td>
<td>54.5</td>
<td>25.0</td>
<td>100.0</td>
<td>5.7</td>
<td>17.8</td>
<td>8.4</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>2.6</td>
<td>58.8</td>
<td>84.6</td>
<td>100.0</td>
<td>5.0</td>
<td>16.2</td>
<td>11.9</td>
</tr>
</tbody>
</table>

Notes: $c, \kappa$ and $\bar{\kappa}$ are such that $\rho = 1 + cN^{-k}T^{-1}$ and $\bar{\rho} = 1 - N^{-\bar{k}}T^{-1}$, where $\rho$ and $\bar{\rho}$ are the autoregressive coefficients in the DGP and GLS detrending procedure, respectively.