TI 2002-002/4

Tinbergen Institute Discussion Paper

How large is Average Economic Growth?

H. Peter Boswijk[1,3]

Philip Hans Franses[2, 3]

[1] *Faculty of Economics and Econometrics, University of Amsterdam,*

[2] *Faculty of Economics, Erasmus University Rotterdam,*

[3] *Tinbergen Institute*

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

How large is average economic growth? Evidence from a robust method

H. Peter Boswijk*
Department of Quantitative Economics
Universiteit van Amsterdam

Philip Hans Franses
Econometric Institute
Erasmus University Rotterdam

December 2001

Abstract

This paper puts forward a method to estimate average economic growth, and its associated confidence bounds, which does not require a formal decision on potential unit root properties. The method is based on the analysis of either difference-stationary or trend-stationary time series models, implementing the robust bootstrapping procedure advocated in Romano and Wolf (2001). Simulation evidence indicates the practical relevance of the method. It is illustrated on quarterly post-war US industrial production.

Key words: Growth, Unit root, Robust testing

1 Introduction

The question of the size of average economic growth seems like a rather trivial one to ask. Yet, time series econometricians know that the answer is far from trivial. Indeed, the answer for the point estimate of average growth hinges upon the time series model employed. Usually, one can choose between a trend-stationary (TS) model and a difference-stationary (DS) model, and often the numerical value of the average growth estimate differs across the two models. Additionally, the associated confidence intervals also depend on the chosen model. Those of the TS model are usually rather narrow, while those of the DS model are rather wide. However, the confidence bounds obtained from the TS model tend to underestimate the true bounds in case the root of the autoregressive model gets close to unity.

As the estimate of average economic growth depends on the model, one may be inclined to make a selection between the models first, and, based on the outcome, then to estimate growth. In this case the selection typically depends on the outcome of a test for a unit root. Unfortunately, these tests have notoriously low power, and hence it is quite likely one ends up with the DS model, while a TS model

*Address for correspondence: Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands; e-mail: peterb@fee.uva.nl. The computer programs used for all calculations in this paper can be obtained from the corresponding author. This paper was prepared for the poster session of the EC-2 (2001) conference in Louvain-la-Neuve.

with a close-to-unity root would have been a better option. Furthermore, the pre-testing aspect of such a procedure tends to complicate the distribution of estimators and associated t-statistics. It seems therefore relevant to try to answer the question in the title while using a method which does not depend (or, at least, not much) on the chosen model. In this paper we put forward such a method.

The analysis is closely related to the work of Canjels and Watson (1997), who consider various point estimators and confidence interval method for the trend slope in a model with a near-unit root. The main difference with their analysis is that we avoid the use of asymptotic critical values, by using the subsampling method recently put forward by Romano and Wolf (2001). Unlike more conventional bootstrap procedures, this subsampling method is asymptotically valid in the presence of a near-unit root, and therefore suitable to obtain robust estimates and confidence intervals for the average economic growth, where the robustness is with respect to the deviation from the unit root.

The outline of this paper is as follows. In Section 2 we discuss the TS and DS models we consider the two associated methods for point and interval estimation of the mean growth rate. We use quarterly seasonally adjusted post World War II US total industrial production as the running example throughout this paper. In Section 3 we provide the asymptotic distribution theory for our procedures, and analyse its implications the empirical examples. In Section 4 we discuss a subsampling method for computing confidence bounds, adapting the elegant approach put forward in Romano and Wolf (2001), together with its application to the industrial production data. Section 5 reports on a simulation experiment which is used to investigate how robust the subsampling method really is, and how reliable it is in smaller samples. In the last section we explore some future research topics.

2 Representation and estimation

Consider a time series y_t which can be described by a first order autoregressive model with trend, that is,

$$\Delta y_t = \beta + \gamma(y_{t-1} - \alpha - \beta[t-1]) + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{1}$$

where Δ denotes the first-order differencing filter, and $\{\varepsilon_t\}$ is an i.i.d. $N(0, \sigma^2)$ process. The starting value y_0 is observed, and is consider fixed. The trend-reversion parameter γ may be zero, such that y_t is a random walk with drift β, or lie in the interval $(-2, 0)$, such that y_t is a trend-stationary AR(1) process with trend slope β. We opt for this representation as it ensures that the focal parameter is β in both cases; when y_t is the natural logarithm of an economic time series Y_t, then β represents the mean growth rate of Y_t. In practice the model will typically be extended to include lagged differences to avoid serial correlation in ε_t. We focus here on the first-order autoregression for clarity, but the results to follow can all be extended to higher-order autoregressions.

To emphasize the matter of concern in this paper, we consider the limiting distribution of the Maximum Likelihood Estimator (MLE) of β and their estimated standard errors in the two cases. For that purpose, consider

$$\Delta y_t = \mu + \tau t + \gamma y_{t-1} + \varepsilon_t, \tag{2}$$

where $\mu = \beta(1+\gamma) - \gamma\alpha$ and $\tau = -\gamma\beta$, and hence

$$\alpha = \begin{cases} [\mu + \tau(1+1/\gamma)]/\gamma, & \text{if } -2 < \gamma < 0, \\ \text{not identified}, & \text{if } \gamma = 0, \end{cases} \tag{3}$$

$$\beta = \begin{cases} -\tau/\gamma, & \text{if } -2 < \gamma < 0, \\ \mu, & \text{if } \gamma = 0. \end{cases} \tag{4}$$

The MLEs of (μ, τ, γ) are obtained by Ordinary Least Squares (OLS) in (2). As $\hat{\gamma} = 0$ with probability zero, the MLE of β is given by $\hat{\beta} = -\hat{\tau}/\hat{\gamma}$, almost surely. We will refer to this as the TS estimator, but we will evaluate its properties also for cases when the true DGP is in fact DS, or TS with a near-unit root. The squared estimated standard error of $\hat{\beta}$ (obtained from the delta method) is then given by

$$\hat{s}^2_{\hat{\beta}} = \frac{1}{\hat{\gamma}^2}(1, \hat{\beta})\hat{V}[\hat{\theta}] \begin{pmatrix} 1 \\ \hat{\beta} \end{pmatrix}, \tag{5}$$

where $\hat{V}[\hat{\theta}]$ is the OLS estimated covariance matrix of $\hat{\theta} = (\hat{\tau}, \hat{\gamma})'$. Denote

$$z_t = \begin{pmatrix} t - n^{-1}\sum_{t=1}^n t \\ y_{t-1} - n^{-1}\sum_{t=1}^n y_{t-1} \end{pmatrix}, \tag{6}$$

and $\theta = (\tau, \gamma)'$, such that $\hat{\theta} - \theta = (\sum_{t=1}^n z_t z_t')^{-1} \sum_{t=1}^n z_t \varepsilon_t$, and $\hat{V}[\hat{\theta}] = \hat{\sigma}^2 (\sum_{t=1}^n z_t z_t')^{-1}$, with $\hat{\sigma}^2$ the OLS residual variance. We also consider an alternative estimate of the standard error of $\hat{\beta}$, viz.

$$\tilde{s}^2_{\hat{\beta}} = \frac{1}{\hat{\gamma}^2}(1, \beta)\hat{V}[\hat{\theta}] \begin{pmatrix} 1 \\ \beta \end{pmatrix}, \tag{7}$$

which uses the value of β under the null hypothesis; this may be used in a t-statistic $(\hat{\beta} - \beta)/\tilde{s}_{\hat{\beta}}$, as an alternative to $(\hat{\beta} - \beta)/\hat{s}_{\hat{\beta}}$. The standard error $\tilde{s}_{\hat{\beta}}$ originates from the fact that the null hypothesis $H_0 : \beta = \beta_0$ may be reformulated, for almost all parameter values (but excluding $\gamma = 0$), as $H_0' : \gamma\beta_0 + \tau = 0$: the t-statistic $(\hat{\beta} - \beta)/\tilde{s}_{\hat{\beta}}$ can be shown to equal the ratio of $(\hat{\gamma}\beta + \hat{\tau})$ and its OLS standard error.

Note that t-statistic based on $\hat{s}_{\hat{\beta}}$ can easily be inverted to obtain a confidence interval, using quantiles of the null distribution of $(\hat{\beta} - \beta)/\hat{s}_{\hat{\beta}}$. From the standard error $\tilde{s}_{\hat{\beta}}$, we may define a confidence interval as the set of β's which are not rejected, using the null distribution of $(\hat{\beta} - \beta)/\tilde{s}_{\hat{\beta}}$; this requires a non-linear search for the bounds of the confidence interval. When γ is close to 0, may expect better finite-sample size behaviour of $\tilde{s}_{\hat{\beta}}$; see Boswijk (1993) for evidence on this in a cointegration context. On the other hand, when γ and τ are equal to zero, β is not identified from $-\tau/\gamma$, which will often lead to unbounded confidence intervals. This can be seen from the fact that

$$\lim_{\beta \to \infty} \frac{\hat{\beta} - \beta}{\tilde{s}_{\hat{\beta}}} = -\lim_{\beta \to -\infty} \frac{\hat{\beta} - \beta}{\tilde{s}_{\hat{\beta}}} = \frac{\hat{\gamma}}{\hat{s}_{\hat{\gamma}}}, \tag{8}$$

where the right-hand side equals the Dickey-Fuller statistic. Therefore, when the Dickey-Fuller test statistic is close to zero, we will not be able to reject any large positive and negative values of β, yielding

an unbounded confidence interval based on $(\hat{\beta}-\beta)/\tilde{s}_{\hat{\beta}}$. In such cases a confidence interval based on $\hat{s}_{\hat{\beta}}$ might be preferable; although it might have a size distortion (leading to undercoverage of the interval), at least it is informative about the possible values of β. Clearly if $\gamma = 0$ and this is known, inference based on the DS model will be optimal, and any inference based on the TS model can be second-best at most.

Before considering the asymptotic properties of the TS-based method, we briefly discuss the DS analysis, obtained by imposing $\gamma = 0$. In that case the model (1) reduces to

$$\Delta y_t = \beta + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{9}$$

leading to $\tilde{\beta} = n^{-1}\sum_{t=1}^{n}\Delta y_t = n^{-1}(y_n - y_0)$, and its OLS standard error $s_{\tilde{\beta}} = \sqrt{\tilde{\sigma}^2/n}$, with $\tilde{\sigma}^2 = (n-1)^{-1}\sum_{t=1}^{n}(\Delta y_t - \tilde{\beta})^2$. This leads to a t-statistic $(\tilde{\beta}-\beta)/s_{\tilde{\beta}}$, which may be inverted to obtain confidence bounds.

To illustrate the consequences for practical analysis of the above results, consider an application to the quarterly observed post WW-II total industrial production index for the United States. The data have been seasonally adjusted and cover the range 1951.1 to 2000.4. All subsequent models include 5 lags to whiten the errors, at least approximately. Suppose we are interested in the annual growth rate of this industrial production series. When we consider the TS model for the natural logarithmic transformed data, we obtain an estimate of this annual growth rate of 3.109 with a standard error of 0.256. Hence, the conventional 95% confidence interval would range from 2.607 to 3.612. If we would adopt the DS model, we impose $\gamma = 0$ and we obtain an estimate of the annual growth rate of 3.524 with associated standard error 0.610, implying a considerably wider 95% confidence interval, ranging from 2.328 to 4.720.

This illustration shows that we not only get substantially different estimates for the annual growth rate (imagine generating 10 year ahead forecasts!) across the TS and DS model, but also that we get rather different confidence bounds. In particular, note that the DS point estimate almost lies outside the TS confidence interval. If we would want to formally choose between the two models, we can implement the familiar augmented Dickey-Fuller test. The test statistic equals -2.224, and hence we cannot reject $\gamma = 0$, and we would select the DS model and the corresponding point estimate and confidence interval for the growth rate.

The problem with such a procedure, however, lies in the notoriously low power of unit root tests against near-integrated alternatives. This means that in practice there is quite a substantial probability of selecting the wrong model. An additional problem is that the distribution of the TS and DS t-statistics may deviate substantially from the standard normal in near-integrated cases; in particular these distributions tend to be sensitive to the deviation from the unit root. The next section makes this sensitivity explicit by studying the behaviour of the TS and DS procedures under the unit root hypothesis, fixed (trend-stationary) alternatives, as well as local alternatives. This analysis will demonstrate the lack of robustness of standard asymptotic inference to small deviations from the chosen model. This will motivate the analysis in Section 4, where we robustify the procedures using the subsampling procedure of Romano and Wolf (2001).

3 Asymptotic theory

We first consider the behaviour of $\hat{\beta}$ and its estimated standard errors for a trend-stationary data-generating process, in which case $-2 < \gamma < 0$.

Theorem 1 *Let y_t be generated by (1) with $-2 < \gamma < 0$. Then*

$$n^{3/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0, 12\sigma^2/\gamma^2), \tag{10}$$

$$n^3 \tilde{s}^2_{\hat{\beta}} \xrightarrow{p} \frac{12\sigma^2}{\gamma^2}, \tag{11}$$

$$n^3 \hat{s}^2_{\hat{\beta}} \xrightarrow{p} \frac{12\sigma^2}{\gamma^2}. \tag{12}$$

Therefore, the t-ratio of $\hat{\beta}$ is asymptotically standard normal, using either one of the alternative standard error estimates.

A proof is given in the Appendix. Theorem 1 replicates the well-known result that conventional asymptotic inference applies in trend-stationary model.

Now we turn to the case where the process is near-integrated, i.e., $\gamma_n = c/n$ for fixed c, including $c = 0$ ($\gamma = 0$). The results of Theorem 2 for $\hat{\beta}$ can be obtained, directly or indirectly, from Canjels and Watson (1997; see also Phillips and Lee, 1996), but we provide a proof of these in the Appendix for convenience. The result for the two standard errors $\hat{s}_{\hat{\beta}}$ and $\tilde{s}_{\hat{\beta}}$ is new (to our knowledge).

Theorem 2 *Let y_t be generated by (1) with $\gamma_n = \dfrac{c}{n}$, where $c \leq 0$ is a constant, and with y_0 fixed. Then*

$$n^{1/2}(\hat{\beta} - \beta) \xrightarrow{d} -\frac{\sigma \left(\int_0^1 G_1^2\right)^{-1} \int_0^1 G_1 dW}{c + \left(\int_0^1 G_2^2\right)^{-1} \int_0^1 G_2 dW} = \sigma\xi, \tag{13}$$

$$n\tilde{s}^2_{\hat{\beta}} \xrightarrow{d} \frac{\sigma^2 \left(\int_0^1 G_1^2\right)^{-1}}{\left[c + \left(\int_0^1 G_2^2\right)^{-1} \int_0^1 G_2 dW\right]^2} = \sigma^2\zeta, \tag{14}$$

$$\begin{aligned} n\hat{s}^2_{\hat{\beta}} &\xrightarrow{d} \sigma^2\zeta + \frac{\sigma^2 \left[\xi^2 \left(\int_0^1 G_2^2\right)^{-1} - 2\xi \left(\int_0^1 (s - \frac{1}{2})(V - \bar{V})\right) \left(\int_0^1 (s - \frac{1}{2})^2 \int_0^1 G_2^2\right)^{-1}\right]}{\left[c + \left(\int_0^1 G_2^2\right)^{-1} \int_0^1 G_2 dW\right]^2} \\ &= \sigma^2\eta. \end{aligned} \tag{15}$$

Here $W(s)$ is a standard Brownian motion process, $V(s) = \int_0^s e^{c(s-u)} dW(u)$ is an Ornstein Uhlenbeck process, and

$$G_1(s) = s - \int_0^1 sV^*(s)ds \left[\int_0^1 V^*(s)V^*(s)'ds\right]^{-1} V^*(s), \tag{16}$$

$$G_2(s) = V(s) - \int_0^1 V(s)f(s)ds \left[\int_0^1 f(s)f(s)'ds\right]^{-1} f(s), \tag{17}$$

where $V^*(s) = (1, V(s))'$ and $f(s) = (1, s)'$; and $\bar{X} = \int_0^1 X(s)ds$ for any process X.

Theorem 2 implies that the t-statistic of $\hat{\beta}$ has as its limiting null distribution either $\xi/\sqrt{\zeta}$ (if $\tilde{s}_{\hat{\beta}}$ is used) or $\xi/\sqrt{\eta}$ (if $\hat{s}_{\hat{\beta}}$ is used), both of which are characterized by a single nuisance parameter c. The corresponding densities, for various values of c, are depicted in Figures 1 and 2.

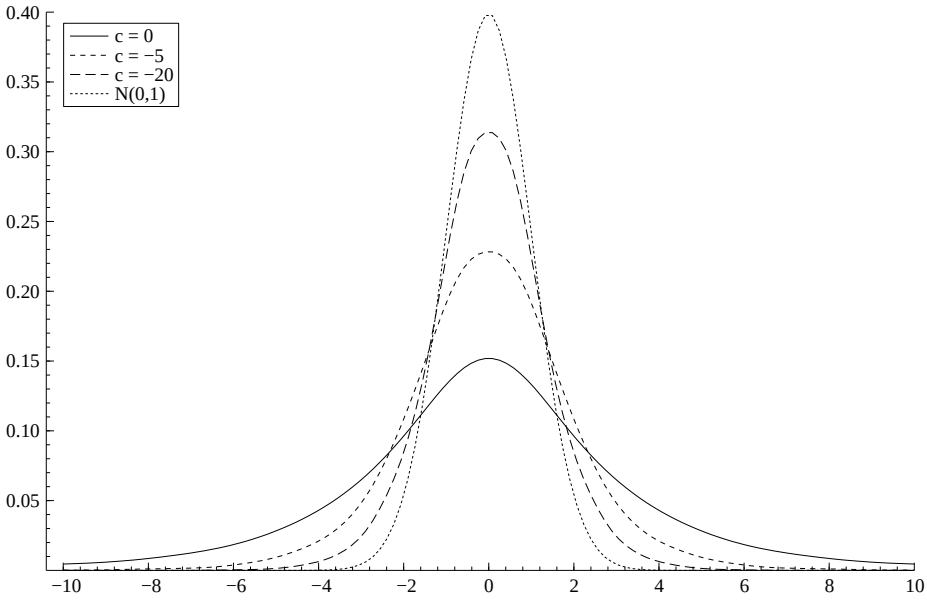

Figure 1. Distribution of the t-statistic using $\hat{s}_{\hat{\beta}}$.

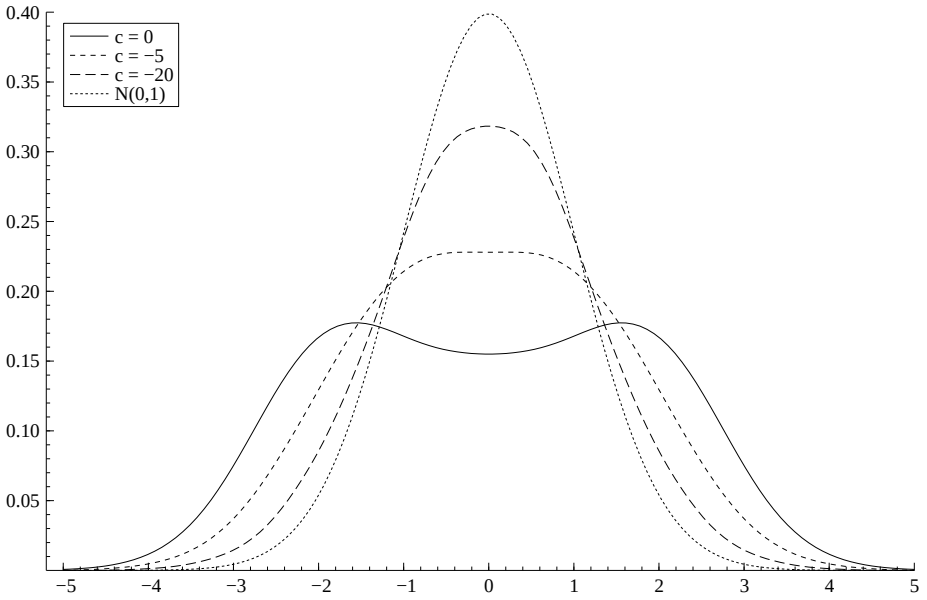

Figure 2. Distribution of the t-statistic using $\tilde{s}_{\hat{\beta}}$.

We observe that the distributions of the t-statistics in Figure 1 can deviate very strongly from the standard normal distribution, and that they are very sensitive to the deviation c from the unit root hypoth-

esis. In the extreme case of a unit root, the t-statistic has a very high dispersion, and the 5% (two-sided) critical values are in the neighbourhood of ± 7.5 instead of the conventional ± 2. When $\tilde{s}_{\hat{\beta}}$ is used, critical values are less dependent on c, and fluctuate from ± 3 ($c = 0$) to ± 2 ($c = -\infty$). Note that in this case the distribution of the t-statistic is bi-modal when $c = 0$, a property inherited from the distribution of the estimated trend coefficient $\hat{\tau}$, see Dickey and Fuller (1981).

Note that Canjels and Watson (1997) also consider the case where y_0 is not fixed, but where the process starts at zero at time $-[\kappa n]$, for some fixed $\kappa > 0$. This leads to an additional nuisance parameter κ; the analysis in Theorem 2, and the inference procedures discussed below, could be extended to cover this case as well.

Consider now the asymptotic behaviour of the estimator $\tilde{\beta}$ and standard error $s_{\tilde{\beta}}$ based on the DS model. The results below for $\tilde{\beta}$ follow as a special case from the results in Canjels and Watson (1997).

Theorem 3 *Let y_t be generated by (1) with $-2 < \gamma < 0$, and with $y_0 \sim N(\alpha, \sigma_y^2)$, where $\sigma_y^2 = \sigma^2/(1-[1+\gamma]^2)$. Then*

$$n(\tilde{\beta} - \beta) \xrightarrow{d} N(0, 2\sigma_y^2), \tag{18}$$

$$ns_{\tilde{\beta}}^2 \xrightarrow{p} -2\gamma\sigma_y^2, \tag{19}$$

so that the t-statistic satisfies $n^{1/2}(\tilde{\beta} - \beta)/s_{\tilde{\beta}} \xrightarrow{d} N(0, -1/\gamma)$.

The fact that the OLS t-statistic converges to zero in this case is related to the MA unit root caused by the DS model. It implies that we may expect serious overcoverage (too wide intervals) of the standard asymptotic confidence interval $\tilde{\beta} \pm 1.96 s_{\tilde{\beta}}$. Note that the asymptotic variance of the normalized t-statistic increases as $\gamma \to 0$; the final theorem investigates the behaviour of the estimator and t-statistic for sequences $\gamma = c/n$.

Theorem 4 *Let y_t be generated by (1) with $\gamma_n = \dfrac{c}{n}$, where $c \leq 0$ is a constant, and with y_0 fixed. Then*

$$n^{1/2}(\tilde{\beta} - \beta) \xrightarrow{d} \sigma V(1) \sim N\left(0, \sigma^2 \operatorname{var}[V(1)]\right), \tag{20}$$

$$ns_{\tilde{\beta}}^2 \xrightarrow{p} \sigma^2, \tag{21}$$

where $V(s)$ is defined in Theorem 2, and

$$\operatorname{var}[V(1)] = \begin{cases} 1, & \text{if } c = 0, \\ \dfrac{1 - e^{2c}}{-2c}, & \text{if } c < 0. \end{cases} \tag{22}$$

The theorem implies that the t-statistic has an asymptotic standard normal distribution only in case of an exact unit root, i.e., when $c = 0$. Since $\operatorname{var}[V(1)] < 1$ for $c < 0$, it follows that we may again expect standard asymptotic confidence intervals $\tilde{\beta} \pm 1.96 s_{\tilde{\beta}}$ to be too wide and hence display overcoverage when $c < 0$.

To illustrate the consequences for practical analysis of the above results, consider again the US industrial production index example. Based on the $1951.1 - 2000.4$ sample ($n = 200$), the estimate of

the trend reversion parameter equals $\hat{\gamma} = -0.033$, which is not significantly different from 0 according to augmented Dickey Fuller statistic of -2.224, as mentioned in the previous section. The corresponding estimate of c is given by $\hat{c} = n\hat{\gamma} = -6.655$; this is not a consistent estimator, in the sense that even as $n \to \infty$, the estimator $\hat{c}$ will vary randomly around c. The asymptotic power of Dickey-Fuller tests is known to be very small (in particular when a trend is included) against local alternatives c in the neighbourhood of -5, which implies that assuming and imposing $c = 0$ would be about equally arbitrary as assuming that, e.g., $c = -5$ (or $c = -6.655$).

These results, and in particular the lack of consistency of $\hat{c}$, imply that we cannot construct asymptotically valid confidence interval based on Theorems 1–4 that are robust to variations in the true c, which motivates the use of subsampling methods in the next section. The results for the TS estimator imply that the TS confidence interval of $3.109 \pm 1.96 \cdot 0.256$ is likely to be far too narrow. Assuming that the true c equals 0 would lead to an asymptotic confidence interval of about $3.109 \pm 7.5 \cdot 0.256 = (1.189, 5.029)$, instead of the $(2.607, 3.612)$ interval corresponding to $c = -\infty$; and if $c = -5$, the appropriate confidence interval would become $3.109 \pm 4 \cdot 0.256 = (2.085, 4.133)$. The confidence intervals based on the TS model using $\tilde{s}_{\hat{\beta}}$ will be unbounded for $c = 0$, -5 or -20; as discussed in the previous section, very large positive or negative values of β will have a t-statistic equal to ± 2.224 (the augmented Dickey-Fuller statistic), which does not lie in the critical region obtained from the densities in Figure 2; hence all these large absolute values of β lie in the confidence region.

Consider finally the implications of Theorems 3 and 4 for the DS confidence interval. Since γ is evidently close to 0, we concentrate on the local-to-unity asymptotics of Theorem 4. When the true value of c equals 0, then the $3.524 \pm 1.96 \cdot 0.610 = (2.328, 4.720)$ interval obtained in the previous section is asymptotically valid. However, when $c = -5$, which might be equally likely, then the variance of the t-statistic equals $(1 - e^{-10})/10 = 0.100$, which means that a better estimate of the true standard error of $\tilde{\beta}$ would be $0.610 \cdot \sqrt{0.100} = 0.193$, leading to a confidence interval of $3.524 \pm 1.96 \cdot 0.193 = (3.146, 3.902)$ which is far more informative. In summary, we see that the asymptotic results, in combination with the fact that c cannot be estimated consistently, leads to a large set of possible confidence intervals, which vary substantially in their width and their location.

4 Confidence intervals based on subsampling

In this section we give a brief discussion of the subsampling method proposed by Romano and Wolf (2001), henceforth RW, applied to our research question. For details and proofs of various results, we refer to the original paper of RW. The basic idea is as follows. Consider the TS t-statistic $(\hat{\beta} - \beta)/\hat{s}_{\hat{\beta}}$ (we will not consider $\tilde{s}_{\hat{\beta}}$ in the remainder of this paper). The construction of a valid confidence interval requires knowledge of the (asymptotic) distribution of this t-statistic, but this distribution depends on γ. However, the distribution may be estimated by the empirical distribution function of t-statistics based on subsamples of length $b < n$ (the block size).

In general, let $\hat{\theta}_{b,n}$ and $\hat{\sigma}_{b,n}$ denote an estimator and a scaling factor based on the tth subsample $\{y_t, y_{t+1}, \ldots, y_{t+b-1}\}$, and let $\hat{\theta}_n$ and $\hat{\sigma}_n$ denote the corresponding statistics based on a full sample of

size n. Furthermore, let τ_n denote a normalizing sequence; such that $\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ has a limiting distribution for all DGPs under consideration; in practice, $\hat{\sigma}_n/\tau_n$ is an estimated standard error of $\hat{\theta}_n$. The empirical distribution of $\{\tau_b(\hat{\theta}_{b,t} - \hat{\theta}_n)/\hat{\sigma}_{b,t}, t = 1, \ldots, n - b + 1\}$ may be used to estimate the distribution of $\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$. The results of RW imply that this estimator is in fact consistent under the following conditions (in addition to some technical conditions discussed in RW):

- $b \to \infty$ as $n \to \infty$, but $b/n \to 0$;
- $a_n(\hat{\theta}_n - \theta)$ and $d_n\hat{\sigma}_n$ both have a limiting distribution, where a_n and d_n are sequences such that $a_b/a_n \to 0$, and $\tau_b/\tau_n \to 0$, with $\tau_n = a_n/d_n$; the limiting distribution of $d_n\hat{\sigma}_n$ may be degenerate, and should have no mass at zero.

Applying the procedure in practice, one has to make a choice about the block size b. Choosing b too small may lead to a bad approximation of the actual distribution, because of small sample-type problems of $\tau_b(\hat{\theta}_{b,t} - \hat{\theta}_n)/\hat{\sigma}_{b,t}$. On the other hand, choosing b too close to n will lead to very little variation in $\{\tau_b(\hat{\theta}_{b,t} - \hat{\theta}_n)/\hat{\sigma}_{b,t}, t = 1, \ldots, n - b + 1\}$, and therefore an underestimation of the true dispersion of the distribution of $\tau_n(\hat{\theta}_n - \beta)/\hat{\sigma}_n$. RW recommend to choose b between $b_{\min} = r_{\min}\sqrt{n}$ and $b_{\max} = r_{\max}\sqrt{n}$, where $r_{\min} \in [0.5, 1]$ and $r_{\max} \in [2, 3]$. The actual choice of b is made by minimizing the local variation of the interval endpoints as a function of b. That is, b should be chosen in a "stable region". The local variation for a choice of b is represented by the so-called volatility index, which is the sum of the moving standard deviations, over $\{b - k, b - k + 1, \ldots, b, \ldots, b + k\}$, of the upper and lower bounds of the confidence interval.

In the present model, Theorems 1 and 2 imply that the conditions of RW are satisfied by $\hat{\beta} = \hat{\theta}_n$ and $\hat{s}_{\hat{\beta}} = \hat{\sigma}_n/\tau_n$, where $\tau_n = n^{1/2}$, and $(a_n, d_n) = (n^{3/2}, n)$ in case $-2 < \gamma < 0$, whereas $(a_n, d_n) = (n^{1/2}, 1)$ when $\gamma = 0$ (note that a_n and d_n are allowed to vary with the nuisance parameter γ, but τ_n is not). Hence asymptotically valid confidence intervals may be obtained by inverting the t-statistic using quantiles from the empirical distribution of the subsample t-statistics. RW distinguish between equal-tailed and symmetric confidence intervals; the former is given by $(\hat{\beta} - q_{0.975}\hat{s}_{\hat{\beta}}, \hat{\beta} - q_{0.025}\hat{s}_{\hat{\beta}})$, whereas the latter is given by $\hat{\beta} \pm q^*_{0.95}\hat{s}_{\hat{\beta}}$, where q_α and q^*_α are the αth quantiles of the subsampling distributions of the t-statistic and its absolute value, respectively. Based on previous experience, RW recommend the symmetric intervals.

It should be emphasized that the consistency result of the subsampling procedure is pointwise, for fixed $\gamma \in (-2, 0]$. It cannot be generalized to uniform consistency, which is easily seen as follows. The local-to-unity sequence $\gamma_n = c/n$ corresponds to a sequence $\gamma^*_b = \gamma_{n_b} = c/n_b = c^*_b/b$, where n_b is the minimal sample size corresponding to a block size b; hence it is the "inverse" of the sequence b_n of block sizes. Now since $b_n/n \to 0$ as $n \to \infty$, it follows that $b/n_b \to 0$ as $b \to \infty$, which implies that $c^*_b = bc/n_b \to 0$. Therefore, under the local-to-unity assumption $\gamma_n = c/n$, the distribution of the subsample t-statistics will converge to the limiting distribution given in Theorem 2 with $c = 0$.

RW show that the subsampling procedure is also able to cope with some mild residual autocorrelation caused by dynamic misspecification. One occasion where such a misspecification occurs is when

the unit root is imposed, whereas in reality $\gamma < 0$; this implies that the disturbance $\varepsilon_t = \Delta y_t - \beta$ in (9) follows an ARMA(1,1) process with an AR root of $1 + \gamma$, and an MA unit root. This might suggest that the same subsampling procedure could also be applied to the estimator $\tilde{\beta}$ and its standard error $s_{\tilde{\beta}}$. However, the different convergence rate of the t-statistics in Theorem 3, caused by the MA unit root, implies that the subsampling procedure will not be consistent in this case. To see this, take τ_n as any sequence, and define $\hat{\sigma}_n = \tau_n s_{\tilde{\beta}}$, so that $\tau_n(\tilde{\beta} - \beta)/\hat{\sigma}_n$ equals the t-statistic. The results of RW require this t-statistic to have a (non-degenerate) limiting distribution for all parameter values, but Theorem 3 implies that when $\gamma < 0$, the t-statistic converges to zero. To avoid this one could redefine $\hat{\sigma}_n$ and τ_n, but that would lead to divergence of the t-statistic under $\gamma = 0$. Essentially we find that the choice of τ_n requires knowledge of the nuisance parameter γ, which demonstrates the lack of robustness of the procedure. A possible solution is to replace $s_{\tilde{\beta}}$ by an estimate of the long-run variance of Δy_t; we do not consider this possibility explicitly in the present paper.

Note that the Theorem 4, in combination with the result indicated earlier that $\gamma_b = c_b/b$ with $c_b \to 0$, suggests that the subsampling procedure might work for local alternatives to the unit root: the distribution of the t-statistic in Theorem 4 when c approaches 0 is the standard normal. Although we do not provide a formal analysis of this, we consider this as a sufficient motivation to study the effectiveness of the subsampling procedure applied to the DS model. Furthermore, we use this method in what follows to see how severe the asymptotic problems really are in practice.

Application of the subsampling procedure to the US production growth data is considered in Figures 3 and 4, which depict the equal-tailed and symmetric confidence intervals as a function of $b \in [40, 100]$ (recall that $n = 200$), together with the associated volatility index ($k = 3$), for the TS and DS models.

Consider first the TS confidence intervals in Figure 3. Both the equal-tailed and the symmetric intervals seem to be fairly stable over different values of b, although the upper bound of the equal-tailed interval seems to decline somewhat with b. Minimizing the volatility index leads to an equal tailed 95% confidence interval of $(1.328, 3.998)$, and a symmetric interval of $(1.343, 4.876)$. It is clear that both intervals are substantially wider than the TS interval based on stationary asymptotics presented in Section 2; the symmetric interval is in fact fairly close to the TS interval based on $c = 0$ derived in Section 3.

In comparison with the TS intervals, the DS confidence intervals in Figure 4 display much more variation with the block size b: the width of both the equal-tailed and the symmetric interval seems to keep on decreasing as b increases, without clearly stabilizing at some level. This is to be expected when the true trend-reversion parameter γ is less than zero: in that case the subsampling distribution of the t-statistic converges to a point mass at zero as b increases, so a decreasing width of the subsampling confidence interval is predicted by theory. The intervals corresponding to the minimal volatility index are $(2.843, 3.904)$ (equal-tailed) and $(2.845, 4.204)$ (symmetric), but the figures indicate that this corresponds to $b = 100$, which indicated that this result might be very sensitive to our choice of the upper bound for the block size, $b_{\max} = 100$.

The symmetric interval is comparable to the DS intervals obtained in Section 3 based on local asymptotics, with c somewhere between 0 and -5. In any case the DS intervals are narrower than

their TS counterparts, and are in particular (supposedly) more informative about the lower bound for the growth rate. In the next section we use a Monte Carlo experiment to investigate how reliable these conclusions from the subsampling method are.

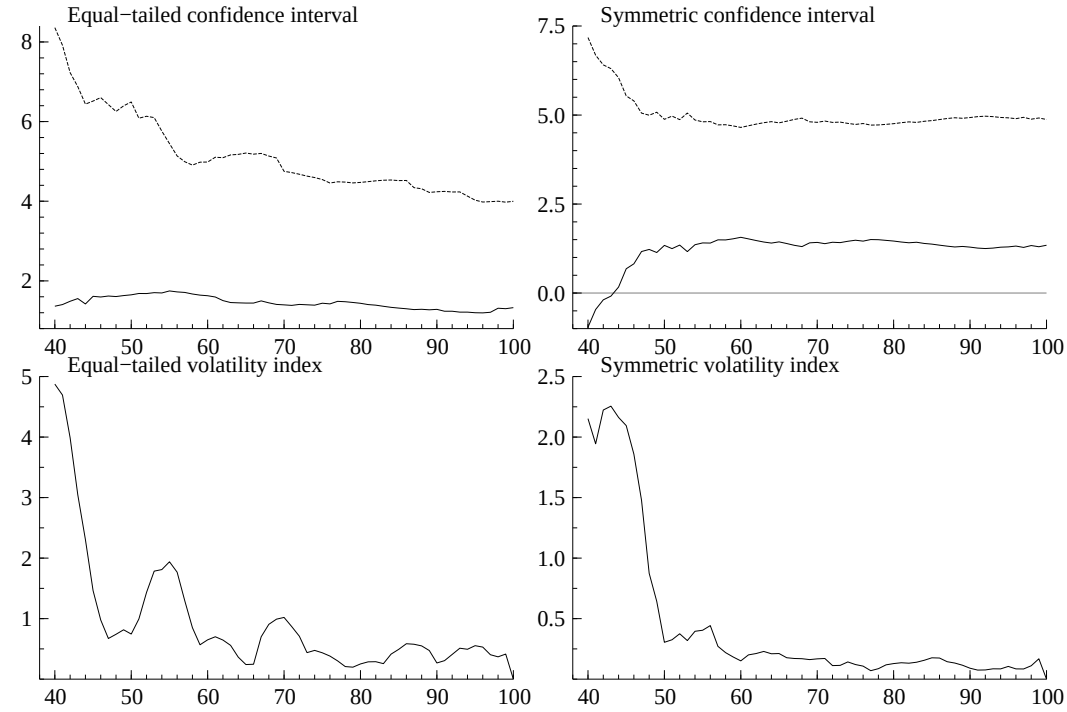

Figure 3. Confidence bounds for β in the trend-stationary model.

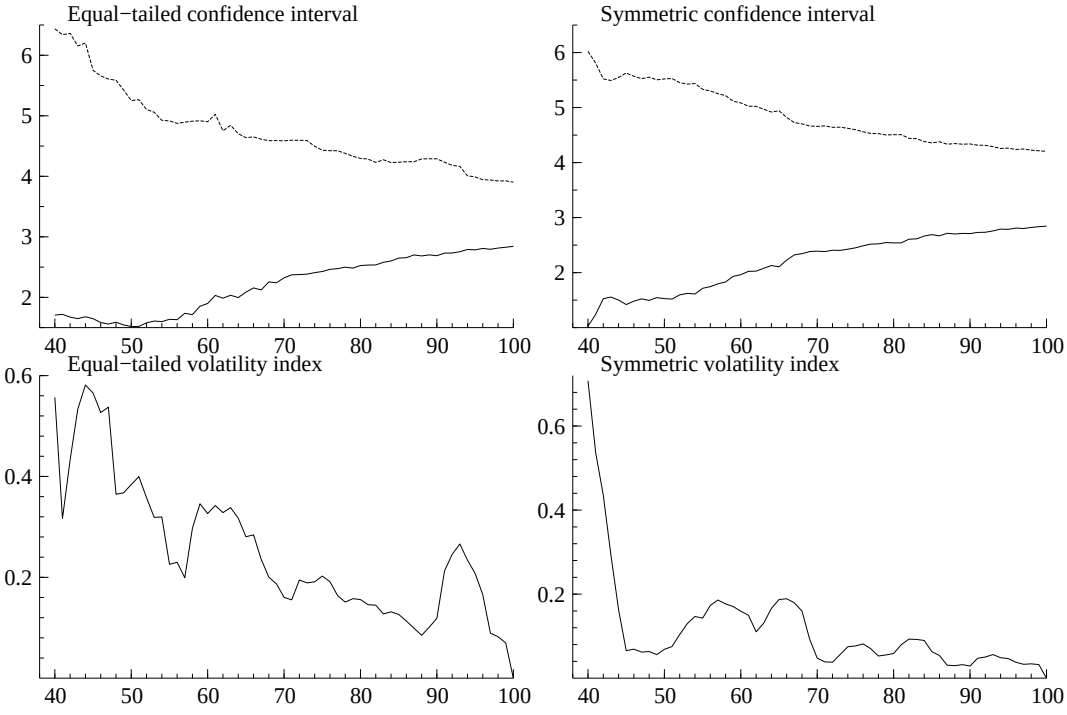

Figure 4. Confidence bounds for β in the difference-stationary model.

5 Simulation evidence

In this section we conduct a small Monte Carlo experiment to investigate the finite sample performance of the subsampling procedure applied to the mean growth rate in the TS and DS model. In the previous section we have seen that the subsampling procedure cannot be expected to be asymptotically valid in the DS model, when the unit root hypothesis is violated; hence we will also investigate how serious these asymptotic problems are in practice.

As the data-generating process, we take the model (1) with $\alpha = \beta = 0$, $\sigma^2 = 1$, $\gamma_n = c/n$, with $n \in \{50, 100, 200, 400\}$ and $c \in \{0, -5, -10, -20\}$. Note that these parameter combinations allow us, to some degree, to investigate the properties for fixed alternatives $\gamma < 0$, e.g. by comparing $(c, n) = (-5, 50)$, $(-10, 100)$ and $(-20, 200)$, which all correspond to the same $\gamma = -0.1$ but different sample sizes.

For each sample, we obtain the TS and DS estimates $(\hat{\beta}, \hat{s}_{\hat{\beta}})$ and $(\tilde{\beta}, s_{\tilde{\beta}})$, and from those, four different confidence intervals:

- the "asymptotic" confidence intervals $\hat{\beta} \pm 1.96\hat{s}_{\hat{\beta}}$ and $\tilde{\beta} \pm 1.96 s_{\tilde{\beta}}$;
- the subsampling confidence interval with $b = b_{\min} = 0.75n^{1/2}$ (except for $n = 50$, where we take $b = 10$);
- the subsampling confidence interval with $b = b_{\max} = 3n^{1/2}$;
- the subsampling confidence interval with an optimal $b \in [b_{\min}, b_{\max}]$, chosen to minimize the volatility index ($k = 2$).

In the Tables below, we report the average (over 1000 replication) coverage rate and width of the various confidence intervals. The nominal coverage rate is 95% in all cases, so a coverage rate substantially less than 0.95 corresponds to too large type I error probabilities. The average width, on the other hand, indicates how informative the confidence intervals are, and hence is related to the power of the procedures.

Table 1 reports the coverage rates of the various implementations of the confidence intervals. We see that the asymptotic TS confidence interval always leads to an undercoverage (overrejection). The subsampling procedure (with optimal block size) does provide a correction for the TS-based confidence interval, but this correction is only fully effective when $c < 0$. For the DS-based procedures, we note that when $c < 0$, we always obtain a coverage rate of 100 per cent, indicating confidence intervals that are wider than necessary. When $c = 0$, the asymptotic DS confidence intervals are clearly valid, but the subsampling procedure with optimal block size seems to lead to some undercoverage, even for large sample sizes.

The average widths of the confidence intervals in Table 2 indicate that the TS confidence intervals are only reasonably informative when $c < -5$ (and $n \geq 100$ when $c = -5$). When $c = 0$, the fact that β is not identified from the TS model leads to extremely large confidence intervals, despite the fact that they still lead to an undercoverage as is clear from Table 1.

Table 1. Coverage rates (%) of TS and DS confidence intervals.

c	n	TS asy	TS min	TS max	TS opt	DS asy	DS min	DS max	DS opt
0	50	56	89	75	77	96	92	79	82
	100	57	93	80	83	95	95	80	83
	200	55	94	85	85	96	96	85	88
	400	55	94	88	89	96	96	89	91
-5	50	74	97	88	88	100	100	99	100
	100	75	99	90	92	100	100	99	100
	200	74	100	95	96	100	100	100	100
	400	75	100	97	97	100	100	100	100
-10	50	83	98	90	92	100	100	100	100
	100	82	100	92	93	100	100	100	100
	200	80	100	96	97	100	100	100	100
	400	82	100	98	98	100	100	100	100
-20	50	90	99	92	93	100	100	100	100
	100	89	100	93	94	100	100	100	100
	200	87	100	97	97	100	100	100	100
	400	88	100	98	99	100	100	100	100

Table 2. Average width of TS and DS confidence intervals.

c	n	TS asy	TS min	TS max	TS opt	DS asy	DS min	DS max	DS opt
0	50	1.688	99.68	119.93	104.31	0.554	0.578	0.370	0.415
	100	0.704	17.27	18.66	18.93	0.393	0.424	0.267	0.302
	200	0.273	1.593	2.309	1.857	0.278	0.303	0.213	0.233
	400	0.411	4.905	12.91	7.596	0.196	0.211	0.166	0.175
-5	50	0.222	0.957	0.586	0.606	0.573	0.522	0.322	0.355
	100	0.218	1.756	2.743	1.914	0.397	0.392	0.230	0.263
	200	0.107	0.418	0.371	0.341	0.280	0.292	0.185	0.205
	400	0.081	0.338	0.338	0.309	0.197	0.205	0.147	0.159
-10	50	0.143	0.479	0.268	0.283	0.589	0.453	0.276	0.300
	100	0.099	0.341	0.207	0.222	0.405	0.364	0.192	0.223
	200	0.066	0.240	0.170	0.171	0.282	0.276	0.157	0.178
	400	0.048	0.181	0.144	0.142	0.198	0.197	0.126	0.141
-20	50	0.085	0.210	0.123	0.127	0.624	0.364	0.232	0.248
	100	0.058	0.168	0.092	0.102	0.416	0.308	0.151	0.175
	200	0.040	0.131	0.081	0.084	0.286	0.248	0.122	0.141
	400	0.028	0.098	0.066	0.068	0.199	0.182	0.099	0.115

The DS procedure seems to have much more stable confidence intervals, which do not vary much with c, but clearly become narrower as n increases. For $c = -10$ the DS and TS have a comparable interval widths, and only when $c = -20$ there is a clear superiority of the TS procedure.

In summary, this Monte Carlo experiment has shown that the subsampling procedure is only very partly effective when applied to the TS model. For a proper coverage rate, the sample size should be large enough and c should be less than zero, so that the desired robustness of the procedure is not fully obtained. The DS-based subsampling procedure on the other hand, despite its theoretical problems mentioned in the previous section, seems to perform much more stable. Although there is a consistent overcoverage when $c < 0$, the width of the confidence intervals is fairly stable across different values of c, and the DS procedure is inferior to the TS procedure only for substantial deviations from the unit root ($c = -20$).

6 Concluding remarks

In this paper we have addressed the problem of making inference on the mean growth rate of a time series when the largest autoregressive root may be close to but is not necessarily equal to one. We have seen that asymptotics only do not provide a solution to this problem, because the non-centrality parameter c, measuring the deviation from the unit root, cannot be estimated consistently. The subsampling procedure of Romano and Wolf (2001) should provide a solution to this problem, at least asymptotically. However, in a small Monte Carlo experiment it has appeared that this method is not fully effective in the trend-stationary model when either the sample size or the deviation from the unit root is small. The same procedure applied to the difference-stationary model seems to be more promising, although there are some problems with its asymptotic validity as indicated in Section 4.

From these infer conclude that the subsampling procedure may be a promising way to deal with the unknown deviation from the unit root, but that we should not limit ourselves to confidence intervals based on either of the two t-statistics considered in this paper. One possible improvement would be to use the DS estimator, but standardized by the square root of the *long-run variance* instead of the usual OLS standard error. Another option might be to use estimators that exploit assumptions about the starting value of the process, and thus might effectively combine information from the DS and TS estimator. We intend to study these extensions in future research.

Further extensions are possible in a multivariate context. In particular, related to Vogelsang and Franses (2001), we may apply the present approach to the question whether, in a panel context, different cross-sectional units (such as countries) have the same growth rate. One might expect by extending the sample size in the cross-sectional direction, smaller time intervals are sufficient to obtain valid inference using the subsampling procedure. These question will also addressed in future research.

Appendix

Proof of Theorem 1. Define

$$D_n = \begin{pmatrix} n^{-3/2} & 0 \\ -\beta n^{-1/2} & n^{-1/2} \end{pmatrix}. \tag{23}$$

Then

$$D_n \sum_{t=1}^{n} z_t z_t' D_n' \xrightarrow{p} \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \sigma_y^2 \end{pmatrix}, \tag{24}$$

$$D_n \sum_{t=1}^{n} z_t \varepsilon_t \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} \frac{1}{12} & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \right). \tag{25}$$

where $\sigma_y^2 = Var(y_t - \beta t) = \sigma^2/(1-[1+\gamma]^2)$. This implies that

$$D_n'^{-1}(\hat{\theta} - \theta) = \begin{pmatrix} n^{3/2}\left[(\hat{\tau}-\tau) + \beta(\hat{\gamma}-\gamma)\right] \\ n^{1/2}(\hat{\gamma}-\gamma) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 12 & 0 \\ 0 & \sigma_y^{-2} \end{pmatrix} \right), \tag{26}$$

and hence

$$\begin{aligned} n^{3/2}(\hat{\beta}-\beta) &= n^{3/2}\left(-\frac{\hat{\tau}}{\hat{\gamma}} - \beta\right) \\ &= -\frac{1}{\hat{\gamma}} n^{3/2}(\hat{\tau} + \beta\hat{\gamma}) \\ &= -\frac{1}{\hat{\gamma}} n^{3/2}\left[(\hat{\tau}-\tau) + \beta(\hat{\gamma}-\gamma)\right] \\ &\xrightarrow{d} N(0, 12\sigma^2/\gamma^2). \end{aligned} \tag{27}$$

Analogously, it can be shown that

$$\begin{aligned} n^3 \hat{s}_{\hat{\beta}}^2 &= \frac{\hat{\sigma}^2}{\hat{\gamma}^2} n^{3/2}(1,\hat{\beta}) D_n' \left[D_n \sum_{t=1}^{n} z_t z_t' D_n' \right]^{-1} n^{3/2} D_n \begin{pmatrix} 1 \\ \hat{\beta} \end{pmatrix} \\ &\xrightarrow{p} \frac{\sigma^2}{\gamma^2}(1,0) \begin{bmatrix} \frac{1}{12} & 0 \\ 0 & \sigma_y^2 \end{bmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{12\sigma^2}{\gamma^2}, \end{aligned} \tag{28}$$

where we have used that $n(\hat{\beta} - \beta) \xrightarrow{p} 0$. The proof of (12) is analogous. □

Proof of Theorem 2. The results use the following functional central limit theorems:

$$n^{-1/2} \sum_{t=1}^{[sn]} \varepsilon_t \xrightarrow{d} B(s) = \sigma W(s), \tag{29}$$

$$n^{-1/2}(y_{[sn]} - \beta[sn]) \xrightarrow{d} U(s) = \sigma V(s) = \sigma \int_0^s e^{-c(s-u)} dW(u), \tag{30}$$

Note that when $c = 0$ (the unit root case), $U(s) = B(s)$, and $V(s) = W(s)$. It is now useful to define

$$D_n = \begin{pmatrix} n^{-3/2} & 0 \\ -\beta n^{-1} & n^{-1} \end{pmatrix}. \tag{31}$$

We then obtain

$$D_n \sum_{t=1}^{n} z_t z_t' D_n' \xrightarrow{d} \int_0^1 \begin{pmatrix} s-\frac{1}{2} \\ U(s)-\bar{U} \end{pmatrix} \begin{pmatrix} s-\frac{1}{2} \\ U(s)-\bar{U} \end{pmatrix}' ds, \tag{32}$$

$$D_n \sum_{t=1}^{n} z_t \varepsilon_t \xrightarrow{d} \int_0^1 \begin{pmatrix} s-\frac{1}{2} \\ U(s)-\bar{U} \end{pmatrix} dB(s). \tag{33}$$

This yields

$$\begin{aligned} D_n'^{-1}(\hat{\theta}-\theta) &= \begin{pmatrix} n^{3/2}\left[(\hat{\tau}-\tau)+\beta(\hat{\gamma}-\gamma)\right] \\ n(\hat{\gamma}-\gamma) \end{pmatrix} \\ &= \begin{pmatrix} -(n\hat{\gamma})\, n^{1/2}(\hat{\beta}-\beta) \\ n\hat{\gamma}-c \end{pmatrix} \\ &\xrightarrow{d} \sigma \left(\int_0^1 FF(s)'ds\right)^{-1} \int_0^1 F(s)dW(s) \\ &= \begin{pmatrix} \sigma \left(\int_0^1 G_1(s)^2 ds\right)^{-1} \int_0^1 G_1(s)dW(s) \\ \left(\int_0^1 G_2(s)^2 ds\right)^{-1} \int_0^1 G_2(s)dW(s) \end{pmatrix}. \end{aligned} \tag{34}$$

where $F(s) = (s-\frac{1}{2}, U(s)-\bar{U})'$. Note that G_2 is V corrected for a constant and trend, and G_1 is a trend corrected for a constant and V. Then

$$n^{1/2}(\hat{\beta}-\beta) = \frac{n^{3/2}\left[(\hat{\tau}-\tau)+\beta(\hat{\gamma}-\gamma)\right]}{c+n(\hat{\gamma}-\gamma)} \xrightarrow{d} -\frac{\sigma\left(\int_0^1 G_1^2\right)^{-1}\int_0^1 G_1 dW}{c+\left(\int_0^1 G_2^2\right)^{-1}\int_0^1 G_2 dW} = \sigma\xi. \tag{35}$$

Furthermore

$$\begin{aligned} n\tilde{s}_{\hat{\beta}}^2 &= \frac{\hat{\sigma}^2}{(n\hat{\gamma})^2} n^{3/2}(1,\beta) D_n' \left[D_n \sum_{t=1}^{n} z_t z_t' D_n'\right]^{-1} n^{3/2} D_n \begin{pmatrix} 1 \\ \beta \end{pmatrix} \\ &\xrightarrow{d} \frac{\sigma^2}{\left[c+\left(\int_0^1 G_2^2\right)^{-1}\int_0^1 G_2 dW\right]^2} (1,0) \left[\int_0^1 FF'\right]^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{aligned} \tag{36}$$

which leads to (14), using the partitioned inverse of $\int_0^1 FF'$, which is

$$\left[\int_0^1 FF'\right]^{-1} = \begin{bmatrix} \left(\int_0^1 G_1^2\right)^{-1} & \sigma \int_0^1 (s-\frac{1}{2})(V-\bar{V}) \left(\sigma^2 \int_0^1 (s-\frac{1}{2})^2 \int_0^1 G_1^2\right)^{-1} \\ * & \left(\sigma^2 \int_0^1 G_1^2\right)^{-1} \end{bmatrix}, \tag{37}$$

where the "$*$" entry follows from symmetry. For $n\hat{s}_{\hat{\beta}}^2$, the result now follows from

$$n^{3/2} D_n \begin{pmatrix} 1 \\ \hat{\beta} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 \\ \sigma\xi \end{pmatrix}. \tag{38}$$

□

Proof of Theorem 3. Let $u_t = y_t - \alpha - \beta t = y_t - E[y_t]$, under the assumptions of the Theorem u_t is a stationary AR(1) process with stationary $N(0, \sigma_y^2)$ distribution. It is easily seen that

$$n(\tilde{\beta} - \beta) = u_n - u_0, \tag{39}$$

which has a $N(0, \sigma_y^2 - 2\operatorname{cov}(u_n, u_0))$ distribution, which converges to $N(0, 2\sigma_y^2)$ since the covariance converges to zero. For $\hat{s}_{\tilde{\beta}}$, we find

$$\begin{aligned} n\hat{s}_{\tilde{\beta}}^2 &= \frac{1}{(n-1)} \sum_{t=1}^{n} (\Delta y_t - \tilde{\beta})^2 \\ &= \frac{1}{(n-1)} \sum_{t=1}^{n} (\Delta u_t)^2 - \frac{n}{(n-1)} (\tilde{\beta} - \beta)^2 \\ &\xrightarrow{p} \operatorname{var}(\Delta u_t) = -2\gamma\sigma_y^2. \end{aligned} \tag{40}$$

□

Proof of Theorem 4. Define the process $U_n(s) = n^{-1/2} u_{[sn]}$ on $D[0,1]$. Then (30) implies that $U_n(\cdot) \xrightarrow{d} U(\cdot) = \sigma V(\cdot)$. Using (39), we have

$$\begin{aligned} n^{1/2}(\tilde{\beta} - \beta) &= n^{-1/2} u_n - n^{-1/2} u_0 \\ &= U_n(1) + \mathrm{o}_p(1) \\ &\xrightarrow{d} \sigma V(1). \end{aligned} \tag{41}$$

As $V(1) = e^c \int_0^1 e^{-cu} dW(u)$, it follows that $V(1) \sim N(0, \operatorname{var}[V(1)])$, with $\operatorname{var}[V(1)]$ as specified in the theorem. For the standard error we have, analogous to (40),

$$\begin{aligned} n s_{\tilde{\beta}}^2 &= \frac{1}{n} \sum_{t=1}^{n} (\Delta u_t)^2 + \mathrm{o}_p(1) \\ &= \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 + \frac{c^2}{n^3} \sum_{t=1}^{n} u_{t-1}^2 + \frac{2c}{n^2} \sum_{t=1}^{n} \varepsilon_t u_{t-1} + \mathrm{o}_p(1) \\ &\xrightarrow{p} \sigma^2, \end{aligned} \tag{42}$$

where we have used $\Delta u_t = (c/n) u_{t-1} + \varepsilon_t$. □

References

Boswijk, H.P. (1993), "On the Formulation of Wald Tests on Long-Run Parameters", *Oxford Bulletin of Economics and Statistics*, 55, 137-144.

Canjels, E. and M.W. Watson (1997), "Estimating Deterministic Trends in the Presence of Serially Correlated Errors", *Review of Economics and Statistics*, 79, 184-200.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, 1057-1072.

Phillips, P.C.B. and C.C. Lee (1996), “Efficiency Gains from Quasi-Differencing under Nonstationarity”, in P.M. Robinson and M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series: Essays in Memory of E.J. Hannan*. New York: Springer Verlag, 300-314.

Romano, J.P. and M. Wolf (2001), “Subsampling Intervals in Autoregressive Models with Linear Time Trend”, *Econometrica*, 69, 1283-1314.

Vogelsang. T.J. (1998), “Trend Function Hypothesis Testing in the Presence of Serial Correlation Parameters”, *Econometrica*, 66, 123-148.

Vogelsang, T.J. and P.H. Franses (2001), “Testing for Common Deterministic Trend Slopes”, Econometric Institute Report 2001-16, Erasmus University Rotterdam.