UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/12

UNIT ROOT TESTS WITH A BREAK IN VARIANCE

by Tae-Hwan Kim, Stephen Leybourne and Paul Newbold

July 2000

DP 00/12

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 00/12

UNIT ROOT TESTS WITH A BREAK IN VARIANCE

by Tae-Hwan Kim, Stephen Leybourne and Paul Newbold

Tae-Hwan Kim is Lecturer, and Stephen Leybourne and Paul Newbold are Professors, School of Economics, University of Nottingham

July 2000

Unit Root Tests With a Break in Variance

Tae-Hwan Kim, Stephen Leybourne and Paul Newbold

Department of Economics
University of Nottingham

June 2000

Abstract

It is shown that an abrupt change in the innovation variance of an integrated process can generate spurious rejections of the unit root null hypothesis in routine applications of Dickey-Fuller tests. We develop and investigate a modified test statistic whose limiting null distribution is invariant to the location and magnitude of a change in innovation variance, and is applicable when the change-point is unknown.

1 Introduction

A few authors have analysed the possibility of breaks in the variance of a time series. For example, Wichern, Miller and Hsu (1976) considered maximum likelihood estimation of an unknown break point in the variance of a first order autoregression, while Hsu (1977) proposed tests for the existence of a break, at an unknown point in time, in the variance of a sequence of independent normal random variables. Inclán (1993) used Bayesian methods to detect multiple breaks in variance in a time series. However, relatively little attention has been paid to the possibility of a break in the innovation variance of an integrated process, and to the impact of such a break on testing the null hypothesis of a unit autoregressive root. An exception is Hamori and Tokihisa (1997). These authors considered Dickey-Fuller tests based only on the regression with no constant and trend, concentrating on the case of an *increase* in innovation variance, reporting a moderate tendency to spuriously reject the unit root hypothesis.

The no constant, no trend model is of very limited practical value, as it implies that, under the alternative hypothesis of trend stationarity, the generating process is known to have mean zero. Unfortunately, the results reported by Hamori and Tokihisa for the simple model turn out to be unreliable predictors, both qualitatively and quantitatively, of what is found when either a constant or a linear trend is incorporated into the Dickey-Fuller regression. In Section 2 of the paper, we analyze the former case in detail and provide simulation evidence of very similar conclusions for the latter. In short, we find quite severe spurious rejections of the unit root null hypothesis when there is a relatively early *decrease* in the innovation variance. [1]

Having demonstrated the phenomenon of spurious rejections by Dickey-Fuller tests in the presence of an innovation variance shift, a result which complements the analysis of a trend shift in Leybourne, Mills and Newbold (1998), the remainder of the paper is devoted to the development of modified Dickey-Fuller tests that allow for a possible change in innovation variance at an unknown point in time. Section 3 of the paper discusses the estimation of the break when one occurs, while Section 4 derives a modified Dickey-Fuller test as an adaptation of a feasible generalised least squares approach. Our test statistic is shown to have an asymptotic null distribution that is invariant to both the location and extent of the break, or indeed to whether

[1] In a footnote, Hamori and Tokihisa appear to suggest, on theoretical grounds, that for their simple model a decrease in variance will lead to *under-rejection* of the null hypothesis. In fact, in simulations, not reported here, of that model based on series of 100 observations, we were unable to confirm that prediction, finding instead a modest tendency to over-reject, particularly for relatively early breaks. Again, however, the phenomenon is far less severe for the simple model than occurs for the more widely used models discussed in Section 2.

or not there is a break. We assess the finite sample size and power of the new test through simulation experiments.

To this point the analysis is concentrated on the driftless random walk DGP

$$y_t = y_{t-1} + \varepsilon_t, \qquad t = 1, ..., T, \tag{1}$$
$$\varepsilon_t = \sigma_t \eta_t \tag{2}$$

where η_t is distributed as $IID(0,1)$ and σ_t^2 is defined by

$$\sigma_t^2 = \sigma_{10}^2 1[t \leq \tau_0 T] + \sigma_{20}^2 1[t > \tau_0 T]. \tag{3}$$

Thus there is a break in the variance of the innovation process ε_t at time $\tau_0 T$, the variance changing from σ_{10}^2 to σ_{20}^2. The appropriate Dickey-Fuller test is then based on a fitted first order autoregression, with constant term. In Section 5 it is shown that the test of Section 4 can be extended to allow for additional serial correlation through the incorporation of lagged changes in the regression, and also for the incorporation of a linear trend, though of course in the latter case different critical values are required.

2 Spurious Rejections in Dickey-Fuller Tests

In this section we analyse the behaviour of the $t-$ratio variant of the Dickey-Fuller test when applied to time series generated by (1)-(3), and based on a fitted first order autoregression. The asymptotic null distribution of the statistic will be derived for the case where the Dickey-Fuller regression includes a constant but no trend. That limiting distribution, which involves both the break fraction τ_0 and the ratio of the innovation standard deviations σ_{20}/σ_{10}, leads to a prediction that the test will spuriously reject the unit root null hypothesis when there is an abrupt *decrease* in innovation variance, most seriously so when the break is relatively early. That prediction is confirmed by simulation evidence which indicates that the problem is potentially quite severe. Parallel simulations for the case where the Dickey-Fuller regression includes a linear trend yield very similar results.

2.1 Asymptotic Distribution of The Dickey-Fuller Statistic

The $t-$ratio variant of Dickey-Fuller statistic, with constant term included, is based on the fitted OLS regression

$$y_t = \hat{\alpha} + \hat{\rho} y_{t-1} + e_t$$

and is given by

$$t_c = \frac{(\hat{\rho} - 1)}{\hat{\sigma}\{\sum_{t=2}^{T}(y_{t-1} - \bar{y})^2\}^{-1/2}} \tag{4}$$

where

$$\hat{\sigma}^2 = T^{-1} \sum_{t=2}^{T} e_t^2$$

The following theorem, proved in the Appendix, gives the asymptotic distribution of t_c under the DGP (1)-(3).

Theorem 1 *Under the DGP (1)-(3),*

$$t_c \Rightarrow \frac{1}{\{\tau_0 + (1-\tau_0)\delta_0^2\}^{1/2}} \frac{A(\delta_0, \tau_0)}{B(\delta_0, \tau_0)^{1/2}} \tag{5}$$

where $\delta_0 = \sigma_{20}/\sigma_{10}$,

$$\begin{aligned} A(\delta_0, \tau_0) &= \frac{1}{2}\{W(1)^2 - 1\} - (\delta_0^2 - 1)\frac{1}{2}\{W(1)^2 - W(\tau_0)^2 - (1-\tau_0)\} \\ &\quad -\delta_0(\delta_0 - 1)W(\tau_0)\{W(1) - W(\tau_0)\} \\ &\quad -[\int_0^1 W(r)dr + (\delta_0 - 1)\{\int_{\tau_0}^1 W(r)dr - (1-\tau_0)W(\tau_0)\}] \\ &\quad .[W(1) + (\delta_0 - 1)\{W(1) - W(\tau_0)\}] \end{aligned}$$

and

$$\begin{aligned} B(\delta_0, \tau_0) &= \int_0^1 W(r)^2 dr + (\delta_0^2 - 1)\int_{\tau_0}^1 W(r)^2 dr \\ &\quad -2\delta_0(\delta_0 - 1)W(\tau_0)\int_{\tau_0}^1 W(r)dr + (\delta_0 - 1)^2(1-\tau_0)W(\tau_0)^2 \\ &\quad -[\int_0^1 W(r)dr + (\delta_0 - 1)\{\int_{\tau_0}^1 W(r)dr - (1-\tau_0)W(\tau_0)\}]^2. \end{aligned}$$

Here $W(r)$ *is a standard Brownian motion process.*

The limit distribution for t_c in (5) is a rather complicated function of the break fraction τ_0 and ratio of standard deviations δ_0, and is not readily interpreted. However, some insight into the effects of a variance break can be obtained by computing the expectations of the terms A and B. Straightforward algebra shows that

$$\begin{aligned} E\{A(\delta_0, \tau_0)\} &= \frac{1}{2}\{\tau_0(\tau_0 - 2) - \delta_0^2(1-\tau_0)^2\}, \\ E\{B(\delta_0, \tau_0)\} &= \frac{1}{6}\{\tau_0^2(3 - 2\tau_0) + \delta_0^2(2\tau_0 + 1)(1-\tau_0)^2\}. \end{aligned}$$

Thus, as a very rough approximation, we have

$$E(t_c) \approx \frac{1}{\{\tau_0 + (1-\tau_0)\delta_0^2\}^{1/2}} \frac{E\{A(\delta_0, \tau_0)\}}{E\{B(\delta_0, \tau_0)\}^{1/2}} \tag{6}$$

The function on the right hand side of (6) does not simplify further, so in Figure 1 we graph the value of the function against the break fraction τ_0, for various values of the standard deviation ratio δ_0. In the case where $\delta_0 = 1.0$ there is, of course, no change in variance, and hence the function takes a constant value over τ_0, equal to -1.225. For $\delta_0 < 1.0$, however, this function is, everywhere in τ_0, more negative than for $\delta_0 = 1.0$, and monotonically so as δ_0 decreases. This effect is most striking for the smaller values of δ_0 when the break point τ_0 is relatively early. In terms of the broad behaviour of the Dickey-Fuller statistic t_c of (4) itself, on the basis of Figure 1, we might expect a decrease in variance under the unit root null to cause the statistic to take a more negative value than would be the case otherwise, and thus lead to the test rejecting the null spuriously when conventional test critical values are applied. The extent to which this happens is investigated by Monte Carlo simulation is the following subsection. In the case where $\delta_0 > 1.0$ the function in (6) takes values greater than -1.225 everywhere in τ_0, but the difference is much less pronounced than for the case when $\delta_0 < 1.0$. This suggests that an increase in variance is very unlikely to lead to severe spurious rejections.

2.2 Monte Carlo Simulation

We conducted simulation experiments for the break in variance DGP (1)-(3) for series of $T = 100$ observations, with η_t generated as standard normal. For τ_0 ranging from 0.01 to 0.99 in steps of 0.01, and $\delta_0 \leq 1.0$ taking the same values as used to generate Figure 1, the empirical size of t_c at the nominal 5%-level was calculated. Each simulation was based on 40,000 replications. The results are shown in Figure 2. Precisely as predicted from our mean approximations, severe spurious rejections occur for low values of τ_0. By way of a check on the generality of this phenomenon, using the same generating model, we also simulated the empirical size of the Dickey-Fuller statistic when a linear trend term as well as constant is included in the fitted OLS regression. The results are shown in Figure 3, where the Dickey-Fuller statistic is denoted t_t. The pattern of spurious rejections is seen to be very similar, both qualitatively and quantitatively, to that of the t_c statistic in Figure 2. For both these tests we conducted additional simulations for the case where $\delta_0 > 1.0$, but we do not report the results as nowhere were the empirical sizes of the tests found to differ substantially from the nominal 5%-level. Again, this is something we might have predicted from the behaviour of the mean approximation, although contrary to that prediction there is a range of τ_0 values in which there is a small tendency for the test to over-reject.

3 Break Estimation

In view of the spurious rejection problem uncovered in the previous section, we now seek a modified Dickey-Fuller test valid in the presence of a possible break in innovation variance. We first consider the estimation of a break point τ_0 assuming that we observe ε_t, and then we relax this assumption later. The basic idea is to transform the structural break in the variance of ε_t into the structural break in the mean of a well-behaved random variable. For such a transformation, the following assumption is imposed.

Assumption 1. $0 < \sigma_x^2 < \infty$ *where* $\sigma_x^2 = var\{\log(\eta_t^2)\}$.

Now define

$$\begin{aligned} Y_t &= \log(\varepsilon_t^2) \\ &= \{\log(\sigma_{10}^2) + \log(\eta_t^2)\}1[t \leq \tau_0 T] + \{\log(\sigma_{20}^2) + \log(\eta_t^2)\}1[t > \tau_0 T] \end{aligned}$$

and

$$\mu(t) = \mu_{10}1[t \leq \tau_0 T] + \mu_{20}1[t > \tau_0 T]$$

with

$$\begin{aligned} \mu_{10} &= E\{\log(\eta_t^2)\} + \log(\sigma_{10}^2), \\ \mu_{20} &= E\{\log(\eta_t^2)\} + \log(\sigma_{20}^2). \end{aligned}$$

Then, the log-transformed variable Y_t follows the process:

$$Y_t = \mu(t) + X_t \tag{7}$$

where $X_t = \log(\eta_t^2) - E\{\log(\eta_t^2)\}$ and X_t is distributed $IID(0, \sigma_x^2)$. Thus, Y_t is a stationary process with a break in its mean at an unknown time $\tau_0 T$, the mean changing form μ_{10} to μ_{20}. While the problem of testing a break at an unknown point has received considerable attention in the literature, rather less attention has been paid to the problem of estimating the unknown break point: procedures that have been proposed and analysed include the MLE method (Picard (1985), Bhattacharya (1987), Fu and Curnow (1990)), the LS method (Bai (1993), Bai and Perron (1998)), the LAD method (Bai (1995)) and the QMLE method (Bai, Lumsdaine and Stock (1998)). For the process in (7), the LS method (Bai (1993), Bai and Perron (1998)) can be directly applied. The LS break estimator is defined by the solution to the LS minimization problem:

$$\begin{aligned} \hat{\tau} &\in \arg\min_{\tau} Q(\tau), \\ Q(\tau) &= \sum_{1}^{\tau T} \{Y_t - \mu_1(\tau)\}^2 + \sum_{\tau T+1}^{T} \{Y_t - \mu_2(\tau)\}^2 \end{aligned}$$

where

$$\begin{aligned}\mu_1(\tau) &= (\tau T)^{-1}\sum_{1}^{\tau T} Y_t,\\ \mu_2(\tau) &= \{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T} Y_t.\end{aligned}$$

It can be shown that the above LS minimization is equivalent to the following maximization problem:

$$\begin{aligned}\hat{\tau} &\in \arg\max_{\tau} |V(\tau)|,\\ V(\tau) &= \{\tau(1-\tau)\}^{1/2}\{\mu_1(\tau)-\mu_2(\tau)\}.\end{aligned}$$

Once we obtain the break estimator $\hat{\tau}$ in this way, the variance estimators are defined by

$$\begin{aligned}\hat{\sigma}_1^2 &= (\hat{\tau} T)^{-1}\sum_{1}^{\hat{\tau} T} \varepsilon_t^2,\\ \hat{\sigma}_2^2 &= \{(1-\hat{\tau})T\}^{-1}\sum_{\hat{\tau} T+1}^{T} \varepsilon_t^2.\end{aligned}$$

The following lemma shows the consistency of these estimators and establishes the convergence rate of the break estimator.

Lemma 1 *Suppose that Assumption 1 holds and* $\tau_0 \in (0,1)$. *Under the DGP(1)-(3),*

$$\begin{aligned}T(\hat{\tau}-\tau_0) = O_p(\lambda_0^{-2}),\\ \hat{\sigma}_1^2-\sigma_{10}^2 = o_p(1),\\ \hat{\sigma}_2^2-\sigma_{20}^2 = o_p(1)\end{aligned}$$

where $\lambda_0 = \log(\sigma_{20}/\sigma_{10})$.

The $T-$consistency result $T(\hat{\tau}-\tau_0) = O_p(\lambda_0^{-2})$ is typical for most break point estimators. In the general case where ε_t is not observable, we use the residuals from the regression $y_t = \hat{\alpha}+\hat{\rho} y_{t-1}+e_t$. Then, the break estimator is defined by:

$$\begin{aligned}\hat{\tau} &\in \arg\max_{\tau} \left|\hat{V}(\tau)\right|,\\ \hat{V}(\tau) &= \{\tau(1-\tau)\}^{1/2}\{\hat{\mu}_1(\tau)-\hat{\mu}_2(\tau)\}\end{aligned}$$

where

$$\begin{aligned}\hat{\mu}_1(\tau) &= (\tau T)^{-1}\sum_{1}^{\tau T}\hat{Y}_t, \\ \hat{\mu}_2(\tau) &= \{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T}\hat{Y}_t\end{aligned}$$

and $\hat{Y}_t = \log(e_t^2)$. The variance estimators are then defined by

$$\begin{aligned}\hat{\sigma}_1^2 &= (\hat{\tau}T)^{-1}\sum_{1}^{\hat{\tau}T}e_t^2, \\ \hat{\sigma}_2^2 &= \{(1-\hat{\tau})T\}^{-1}\sum_{\hat{\tau}T+1}^{T}e_t^2\end{aligned}$$

The following lemma shows that the difference between the two objective functions (one based on ε_t and the other based on e_t) tends to vanish in probability for a fixed τ.

Lemma 2 *Suppose that Assumption 1 holds. Under the DGP (1)-(3),*

$$\left|V(\tau) - \hat{V}(\tau)\right| = o_p(1)$$

for a fixed $\tau \in (0,1)$.

Alternatively, the MLE method by Bhattacharya (1987) can be used to obtain the ML estimators for $\sigma_{10}^2, \sigma_{20}^2$ and τ_0 assuming that ε_t is normally distributed. Given that the normality assumption on ε_t is very restrictive, Bai, Lumsdaine and Stock (1998) proposed the QMLE method for estimation of a break point in which the underlying distribution is not necessarily normal. It can be shown that both the MLE and QMLE methods deliver the same rate of convergence as in lemma 1. While the QMLE method is theoretically appealing, the log-transformation method has substantial numerical advantages in practical implementation. Given the encouraging theoretical results of lemmas 1 and 2, and the motivation to keep the methodology as straightforward as possible to encourage practical use, in the next section we will assess through simulation experiments the performance of a modified Dickey-Fuller test in which break estimation through the log-transformation method is incorporated.

4 Tests based on GLS

We now turn to a modification of the Dickey-Fuller test. Perhaps a natural approach is through feasible generalised least squares. However, we begin by showing that, even if the break date and two innovation variances are known, GLS generates a test statistic whose limiting null distribution depends on nuisance parameters. We next show how the procedure can be amended to avoid this problem. Finally, a feasible procedure, employing the estimators of Section 3 is proposed and assessed.

4.1 Standard GLS

In this subsection we demonstrate that a direct application of the standard GLS does not deliver a desirable solution even when we know all the nuisance parameters σ_{10}, σ_{20} and τ_0. This is because the asymptotic distribution of the $t-$statistic from the standard GLS regression still depends on δ_0, the ratio of the two standard deviations, and the break point τ_0. The application of standard GLS results from

$$\begin{aligned} \sigma_{10}^{-1} y_t &= \sigma_{10}^{-1} y_{t-1} + \sigma_{10}^{-1}\varepsilon_t \quad t = 1, ..., \tau_0 T, \\ \sigma_{20}^{-1} y_t &= \sigma_{20}^{-1} y_{t-1} + \sigma_{20}^{-1}\varepsilon_t \quad t = \tau_0 T + 1, ..., T \end{aligned}$$

which can be rewritten as

$$\tilde{y}_t = \dot{y}_{t-1} + \eta_t \quad t = 1, ..., T$$

where

$$\begin{aligned} \tilde{y}_t &= \sigma_{10}^{-1} y_t 1[t \leq \tau_0 T] + \sigma_{20}^{-1} y_t 1[t > \tau_0 T], \\ \dot{y}_{t-1} &= \sigma_{10}^{-1} y_{t-1} 1[t \leq \tau_0 T] + \sigma_{20}^{-1} y_{t-1} 1[t > \tau_0 T]. \end{aligned} \tag{8}$$

The $t-$statistic is based on the GLS regression

$$\tilde{y}_t = \hat{\alpha}_G + \hat{\rho}_G \dot{y}_{t-1} + e_{Gt}$$

and is given by

$$\begin{aligned} t_{cG} &= \frac{(\widehat{\rho}_G - 1)}{\hat{\sigma}_G \{\sum_{t=2}^T (\dot{y}_{t-1} - \bar{y})^2\}^{-1/2}} \\ &= \frac{T^{-1}\sum_{t=2}^T (\dot{y}_{t-1} - \bar{y})\eta_t}{\hat{\sigma}_G \{T^{-2}\sum_{t=2}^T (\dot{y}_{t-1} - \bar{y})^2\}^{1/2}} \end{aligned}$$

where $\bar{y}$ is the sample mean of $\{\dot{y}_t\}$ and $\hat{\sigma}_G^2 = T^{-1}\sum_{t=2}^T e_{Gt}^2$. The following lemma shows the dependence of the asymptotic null distribution of t_{cG} on δ_0 and τ_0.

Lemma 3 *Under the DGP (1)-(3),*

$$t_{cG} \Rightarrow \frac{A_G(\delta_0, \tau_0)}{B_G(\delta_0, \tau_0)^{1/2}}$$

where

$$\begin{aligned} A_G(\delta_0, \tau_0) &= (1/2)\{W(1)^2 - 1\} - \delta_0^{-1}(\delta_0 - 1)W(\tau_0)\{W(1) - W(\tau_0)\} \\ &\quad -\{\int_0^1 W(r)dr - \delta_0^{-1}(\delta_0 - 1)(1 - \tau_0)W(\tau_0)\}W(1) \end{aligned}$$

and

$$\begin{aligned} B_G(\delta_0, \tau_0) &= \int_0^1 W(r)^2 dr - 2\delta_0^{-1}(\delta_0 - 1)W(\tau_0)\int_0^1 W(r)dr \\ &\quad +\delta_0^{-2}(\delta_0 - 1)^2(1 - \tau_0)W(\tau_0)^2 \\ &\quad -\{\int_0^1 W(r)dr - \delta_0^{-1}(\delta_0 - 1)(1 - \tau_0)W(\tau_0)\}^2. \end{aligned}$$

Here $W(r)$ is a standard Brownian motion process.

Once we express $\tilde{y}_t$ and $\dot{y}_t$ in the form

$$\begin{aligned} \tilde{y}_t &= w_t 1[t \le \tau_0 T] + \{w_t - \delta_0^{-1}(\delta_0 - 1)w_{\tau_0 T}\}1[t > \tau_0 T], \\ \dot{y}_t &= w_t 1[t < \tau_0 T] + \{w_t - \delta_0^{-1}(\delta_0 - 1)w_{\tau_0 T}\}1[t \ge \tau_0 T] \end{aligned}$$

where $w_t = \sum_{i=1}^{t} \eta_i$, the proof of the lemma is quite straightforward and is omitted. The dependence of the asymptotic distribution of t_{cG} on δ_0 is caused by the presence of the term $\delta_0^{-1}(\delta_0 - 1)w_{\tau_0 T}$ in the second regime. This term does not change with t and is not be removed by including a constant in the GLS regression (i.e. by subtracting the sample mean based on the whole sample) because it appears only in the second regime. This observation leads us to propose a modification of the standard GLS approach in which we compute two sub-sample means and subtract the corresponding sub-sample mean from the GLS-transformed variable.

4.2 Modified GLS

Based on the previous subsection, we propose a new method, Modified GLS (MGLS) to derive a test statistic whose asymptotic null distribution is free of all nuisance parameters (σ_{10}, σ_{20} and τ_0). This procedure is based on the assumption that we know the true values of the nuisance parameters. Once we develop a relevant statistic, we will replace the parameters with their consistent estimators. We define

$$\breve{y}_t = (\tilde{y}_t - \bar{y}_{(1)})1[t \le \tau_0 T] + (\tilde{y}_t - \bar{y}_{(2)})1[t > \tau_0 T]$$

where $\tilde{y}_t$ is defined in (8) and

$$\begin{aligned}\bar{y}_{(1)} &= (\tau_0 T)^{-1}\sum_{1}^{\tau_0 T}\tilde{y}_t,\\ \bar{y}_{(2)} &= \{(1-\tau_0)T\}^{-1}\sum_{\tau_0 T+1}^{T}\tilde{y}_t.\end{aligned}$$

The regression of $\breve{y}_t$ on $\breve{y}_{t-1}$ (without a constant) is given by

$$\breve{y}_t = \hat{\rho}_M \breve{y}_{t-1} + e_{Mt}$$

and the $t-$statistic by

$$\begin{aligned}t_{cm} &= \frac{(\widehat{\rho}_M - 1)}{\hat{\sigma}_M(\sum_{t=2}^{T}\breve{y}_{t-1}^2)^{-1/2}}\\ &= \frac{T^{-1}\sum_{2}^{\tau_0 T}\breve{y}_{t-1}\triangle\breve{y}_t + T^{-1}\sum_{\tau_0 T+1}^{T}\breve{y}_{t-1}\triangle\breve{y}_t}{\hat{\sigma}_M(T^{-2}\sum_{2}^{\tau_0 T}\breve{y}_{t-1}^2 + T^{-2}\sum_{\tau_0 T+1}^{T}\breve{y}_{t-1}^2)^{1/2}}\end{aligned}$$

where $\hat{\sigma}_M^2 = T^{-1}\sum_{t=2}^{T} e_{Mt}^2$. Note that we can express $\breve{y}_t$ as

$$\breve{y}_t = (w_t - \bar{w}_{(1)})1[t \le \tau_0 T] + (w_t - \bar{w}_{(2)})1[t > \tau_0 T]$$

where w_t is defined as above and

$$\begin{aligned}\bar{w}_{(1)} &= (\tau_0 T)^{-1}\sum_{2}^{\tau_0 T} w_t,\\ \bar{w}_{(2)} &= \{(1-\tau_0)T\}^{-1}\sum_{\tau_0 T+1}^{T} w_t\end{aligned}$$

Then, it is straightforward to show that $\hat{\sigma}_M \xrightarrow{p} 1$ and

$$t_{cm} \Rightarrow \frac{\tau_0 G_1 + (1-\tau_0)G_2}{\{\tau_0^2 H_1 + (1-\tau_0)^2 H_2\}^{1/2}}$$

where

$$\begin{aligned}G_1 &= (1/2)\{W_1(1)^2 - 1\} - \int_0^1 W_1(r)dr W_1(1),\\ G_2 &= (1/2)\{W_2(1)^2 - 1\} - \int_0^1 W_2(r)dr W_2(1),\\ H_1 &= \int_0^1 W_1(r)^2 dr - \{\int_0^1 W_1(r)dr\}^2,\\ H_2 &= \int_0^1 W_2(r)^2 dr - \{\int_0^1 W_2(r)dr\}^2\end{aligned}$$

and $W_1(r)$ and $W_2(r)$ are two independent standard Brownian motion processes. The limiting distribution of t_{cm} depends on only the one nuisance parameter τ_0 in a symmetrical way, which leads us to propose the following MGLS test statistic:

$$t_{cM} = \frac{\tau_0^{-1} \sum_2^{\tau_0 T} \breve{y}_{t-1} \triangle \breve{y}_t + (1-\tau_0)^{-1} \sum_{\tau_0 T+1}^{T} \breve{y}_{t-1} \triangle \breve{y}_t}{\{\tau_0^{-2} \sum_2^{\tau_0 T} \breve{y}_{t-1}^2 + (1-\tau_0)^{-2} \sum_{\tau_0 T+1}^{T} \breve{y}_{t-1}^2\}^{1/2}}$$

Since $\hat{\sigma}_M \xrightarrow{p} 1$, there is no need to use $\hat{\sigma}_M$ in the MGLS statistic. The following lemma shows that the asymptotic null distribution of the MGLS statistic is free of all nuisance parameters.

Lemma 4 *Under the DGP (1)-(3),*

$$t_{cM} \Rightarrow \frac{G_1 + G_2}{(H_1 + H_2)^{1/2}}.$$

4.3 Feasible Modified GLS

The MGLS statistic t_{cM} is not computable as it involves the parameters σ_{10}, σ_{20} and τ_0. We now propose a computable test statistic which has the same asymptotic null distribution. We assume that we are given estimators $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ and $\hat{\tau}$ with the properties $\hat{\sigma}_1^2 - \sigma_{10}^2 = o_p(1)$, $\hat{\sigma}_2^2 - \sigma_{20}^2 = o_p(1)$ and $T(\hat{\tau} - \tau_0) = O_p(\lambda_0^{-2})$. These estimators can be obtained by either the log-transformation method described in Section 3 or the QMLE method proposed by Bai, Lumsdaine and Stock (1998).

The Feasible MGLS (FMGLS) statistic proposed here is identical to the MGLS statistic except that we use $\hat{\sigma}_1, \hat{\sigma}_2$ and $\hat{\tau}$ in place of σ_{10}, σ_{20} and τ_0. Define now

$$\tilde{y}_t = \hat{\sigma}_1^{-1} y_t 1[t \leq \hat{\tau} T] + \hat{\sigma}_2^{-1} y_t 1[t > \hat{\tau} T]$$

and

$$\breve{y}_t = (\tilde{y}_t - \bar{y}_{(1)}) 1[t \leq \hat{\tau} T] + (\tilde{y}_t - \bar{y}_{(2)}) 1[t > \hat{\tau} T]$$

where

$$\begin{aligned} \bar{y}_{(1)} &= (\hat{\tau} T)^{-1} \sum_1^{\hat{\tau} T} \tilde{y}_t, \\ \bar{y}_{(2)} &= \{(1-\hat{\tau})T\}^{-1} \sum_{\hat{\tau} T+1}^{T} \tilde{y}_t. \end{aligned}$$

The regression of $\breve{y}_t$ on $\breve{y}_{t-1}$ (without a constant) is given by

$$\breve{y}_t = \hat{\rho}_F \breve{y}_{t-1} + e_{Ft}$$

and the proposed FMGLS $t-$statistic is defined by

$$t_{cF} = \frac{\hat{\tau}^{-1}\sum_{1}^{\hat{\tau}T}\breve{y}_{t-1}\triangle\breve{y}_t + (1-\hat{\tau})^{-1}\sum_{\hat{\tau}T+1}^{T}\breve{y}_{t-1}\triangle\breve{y}_t}{\{\hat{\tau}^{-2}\sum_{1}^{\hat{\tau}T}\breve{y}_{t-1}^2 + (1-\hat{\tau})^{-2}\sum_{\hat{\tau}T+1}^{T}\breve{y}_{t-1}^2\}^{1/2}}.$$

We now show that the FMGLS $t-$statistic has the same asymptotic distribution as the MGLS $t-$statistic.

Theorem 2 *Under the DGP (1)-(3),*

$$t_{cF} \Rightarrow \frac{G_1 + G_2}{(H_1 + H_2)^{1/2}}.$$

Finally, we consider the behavior of the FMGLS statistic when there is no break in variance in the DGP. Suppose that y_t is generated by

$$y_t = y_{t-1} + \varepsilon_t, \qquad t = 1, ..., T, \tag{9}$$

where ε_t is distributed $IID(0, \sigma_0^2)$. Nunes, Kuan and Newbold (1995) and Bai (1998) showed that the break estimator $\hat{\tau}$ obtained by searching the unit interval $[0,1]$ converges to 0 or 1 in probability. In both reality and theory, we estimate the break point over an interior $[\tau_1, \tau_2]$ of the unit interval where $0 < \tau_1 < \tau_2 < 1$. Nunes, Kuan and Newbold (1995) and Bai (1998) also showed in this case that the break estimator $\hat{\tau}$ converges in distribution to a well behaved random variable which is a solution to some quadratic maximization problem. These results are directly applicable here, as the log-transformation method of Section 3 converts the problem to one of estimating a break in mean. The following theorem shows that the FMGLS statistic has the same limiting null distribution even when there is no structural break in variance.

Theorem 3 *Under the DGP (9),*

$$t_{cF} \Rightarrow \frac{G_1 + G_2}{(H_1 + H_2)^{1/2}}.$$

4.4 Monte Carlo Simulation

We obtained null critical values for the test statistic t_{cF} via simulation. Data were generated from the DGP (1)-(3) with standard normal η_t. In the simulations we set $\delta_0 = 1.0$, so that no break in variance occurs. When estimating τ_0 we implement the search procedure between sample observations $0.05T$ and $0.95T$, using the log-transformation method of Section 3.

Critical values for different T, based on 40,000 replications, are shown in Table 1.[2] The critical values settle down quite quickly with increasing

[2] The entries under ∞ are actually based on $T = 10,000$.

T. In the case where a break in variance does take place the finite sample distribution of t_{cF} will depend, to some extent, on δ_0 and τ_0. Thus, finite sample critical values computed assuming $\delta_0 = 1.0$ are not really appropriate in the case where $\delta_0 \neq 1.0$. In view of this, we suggest that the test should always be carried out using the asymptotic critical values.

Tables 2 and 3 examine the finite sample size of t_{cF} for different δ_0 and τ_0 using the asymptotic 5% critical value. These simulations are based on 5,000 replications. The values of δ_0 chosen match those used in Figures 1 and 2. For $T = 100$, the test t_{cF} is over-sized for small values of δ_0 and τ_0, relative to the case where no break occurs. For example, with $\delta_0 = 0.25$ and $\tau_0 = 0.15$ its actual size is 7.1%. However, this compares very favourably with the standard Dickey-Fuller test, which, from Figure 2, has a size of about 43%. Increasing the sample size to $T = 200$ improves the size of t_{cF}, as is clear from Table 3. Given that both the location and amount of any break in innovation variance are unknown, and given further the severe size distortions that can occur from routine application of the standard Dickey-Fuller test in the presence of a break, the proposed test appears to exhibit very satisfactory size properties. In this sense at least, our choice of the relatively simple log-transformation method to estimate the break point seems to be vindicated.

Tables 4 and 5 examine the power of t_{cF} to reject a false unit root null in samples of 100 and 200 observations when the true DGP is a first order autoregression with parameter ρ. The test power depends quite heavily on δ_0 and τ_0 in the smaller sample. This is partly due to the fact that test size depends on these parameters also. However, it is quite clear that test power everywhere increases with distance away from the unit root null. Moreover, on comparing Tables 4 and 5, the test appears to be consistent. It is of interest to compare the power of t_{cF} to that of the standard Dickey-Fuller test computed assuming no break in variance. With $\rho = 0.9, 0.8$, our simulations showed the power of the Dickey-Fuller test to be $0.308, 0.862$ for For $T = 100$ and $0.846, 1.000$ for For $T = 200$. Thus, it is clear that t_{cF} generally does not match the Dickey-Fuller test in terms of power. This is of course to be expected as t_{cF} is not constructed with optimality foremost in mind, but instead size robustness under an unknown break in variance. Its gain over the Dickey-Fuller test in terms of test size appears well worth the sacrifice in test power - that is, in many circumstances, it may be worth accepting diminished power in exchange for reliable size and the avoidance of spurious rejections.

5 Extensions

To allow our test statistic t_{cF} to be generalized to account for additional serial correlation, and possibly linear trends, it proves convenient to express it in a different form. First note that t_{cF} can be written as

$$t_{cF} = \frac{\hat{\tau}^{-1}G_{1T} + (1-\hat{\tau})^{-1}G_{2T}}{\{\hat{\tau}^{-2}H_{1T} + (1-\hat{\tau})^{-2}H_{2T}\}^{1/2}}$$

where

$$\begin{aligned} G_{1T} &= \sum_{1}^{\hat{\tau}T} \breve{y}_{t-1}\triangle\breve{y}_t, \qquad H_{1T} = \sum_{1}^{\hat{\tau}T} \breve{y}_{t-1}^2, \\ G_{2T} &= \sum_{\hat{\tau}T+1}^{T} \breve{y}_{t-1}\triangle\breve{y}_t, \qquad H_{2T} = \sum_{\hat{\tau}T+1}^{T} \breve{y}_{t-1}^2. \end{aligned}$$

Now define

$$\begin{aligned} \tilde{y}_{1t} &= \tilde{y}_t, \qquad t = 1, ..., \hat{\tau}T, \\ \tilde{y}_{2t} &= \tilde{y}_{t+\hat{\tau}T}, \qquad t = 1, ..., (1-\hat{\tau})T \end{aligned}$$

and consider the two fitted OLS regressions

$$\tilde{y}_{1t} = \hat{\alpha}_1 + \hat{\rho}_1\tilde{y}_{1t-1} + e_{1t}, \qquad t = 1, ..., \hat{\tau}T, \tag{10}$$

$$\tilde{y}_{2t} = \hat{\alpha}_2 + \hat{\rho}_2\tilde{y}_{2t-1} + e_{2t}, \qquad t = 1, ..., (1-\hat{\tau})T. \tag{11}$$

Then it is straightforward to show that

$$\begin{aligned} G_{1T} &= \frac{(\hat{\rho}_1 - 1)}{s_1^{-2}\hat{V}(\hat{\rho}_1)}, \qquad H_{1T} = \frac{1}{s_1^{-2}\hat{V}(\hat{\rho}_1)}, \\ G_{2T} &= \frac{(\hat{\rho}_2 - 1)}{s_2^{-2}\hat{V}(\hat{\rho}_2)}, \qquad H_{2T} = \frac{1}{s_2^{-2}\hat{V}(\hat{\rho}_2)} \end{aligned}$$

where $\hat{V}(\hat{\rho}_1)$ is the estimated variance of $\hat{\rho}_1$ and s_1^2 is the estimated residual variance from the regression (10), and $\hat{V}(\hat{\rho}_2)$ and s_2^2 are the analogous quantities from the regression (11). This means we can express t_{cF} as

$$t_{cF} = \frac{\hat{\tau}^{-1}\frac{(\hat{\rho}_1-1)}{s_1^{-2}\hat{V}(\hat{\rho}_1)} + (1-\hat{\tau})^{-1}\frac{(\hat{\rho}_2-1)}{s_2^{-2}\hat{V}(\hat{\rho}_2)}}{\{\hat{\tau}^{-2}\frac{1}{s_1^{-2}\hat{V}(\hat{\rho}_1)} + (1-\hat{\tau})^{-2}\frac{1}{s_2^{-2}\hat{V}(\hat{\rho}_2)}\}^{1/2}} \tag{12}$$

which, from a computation aspect, is a rather more transparent and convenient representation.

5.1 Additional Serial Correlation

Suppose that, instead of (1), y_t is generated by the process

$$\Delta y_t = \sum_{i=1}^{k} \phi_i \Delta y_{t-i} + \varepsilon_t \tag{13}$$

Then, the appropriate test statistic takes the same form as (12) except that $\hat{\tau}$, $\hat{\sigma}_1$and $\hat{\sigma}_2$ are calculated from the residuals of the augmented regression

$$y_t = \hat{\alpha} + \hat{\rho} y_{t-1} + \sum_{i=1}^{k} \hat{\phi}_i \Delta y_{t-i} + e_t \tag{14}$$

and $\hat{V}(\hat{\rho}_1)$, s_1^2 and $\hat{V}(\hat{\rho}_2)$, s_2^2 are calculated, respectively, from the two augmented OLS regressions

$$\tilde{y}_{1t} = \hat{\alpha}_1 + \hat{\rho}_1 \tilde{y}_{1t-1} + \sum_{i=1}^{k} \hat{\phi}_{1i} \Delta \tilde{y}_{1t-i} + e_{1t}, \tag{15}$$

$$\tilde{y}_{2t} = \hat{\alpha}_2 + \hat{\rho}_2 \tilde{y}_{2t-1} + \sum_{i=1}^{k} \hat{\phi}_{2i} \Delta \tilde{y}_{2t-i} + e_{2t} \tag{16}$$

The test statistic t_{cF} in (12), based on these newly calculated quantities, continues to have the asymptotic null distribution $(G_1 + G_2)/(H_1 + H_2)^{1/2}$. The proof is straightforward, but lengthy and follows along the lines of the proof of ADF test limiting distribution in the standard case. Tables 6 and 7 give simulations of the finite sample size of t_{cF} for data generated from (13) in the case where $k = 1$ and $\phi_1 = 0.5$. The size results are pretty much comparable to those of the basic test given in Tables 2 and 3 and similar comments therefore apply.

5.2 Linear Trends

As it stands, the test t_{cF} is invariant to transformations of the form $y_t \to c_1 + y_t$ for $\rho \leq 1$. However, in practice, we often require invariance to the transformation $y_t \to c_1 + c_2 t + y_t$. We therefore consider a second test statistic, which we denote t_{tF}. This test is constructed in an entirely analogous manner to t_{cF}, except that we include linear trend terms in the regressions (14), (15) and (16). The asymptotic distribution of t_{tF} has the same form as that of t_{cF}, but whereas that of t_{cF} is written in terms demeaned Brownian motion processes, that of t_{tF} replaces these with demeaned and detrended Brownian motions. The proof of this result is quite straightforward and is not presented here. We have simulated null critical values for t_{tF} and these are provided in Table 8.

References

[1] Bai, J. (1993), "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453-472.

[2] Bai, J. (1995), "Least Absolute Deviation Estimation of a Shift," *Econometric Theory*, 11, 403-436.

[3] Bai, J. (1998), "A Note on Spurious Break," *Econometric Theory*, 14, 663-669.

[4] Bai, J., R. Lumsdaine and J. Stock (1998), "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395-432.

[5] Bai, J. and P. Perron (1998), "Testing for and Estimation of Multiple Structural Changes," *Econometrica*, 66, 47-79.

[6] Bhattacharya, P.K. (1987) "Maximum Likelihood Estimation of a Change-Point in the Distribution of Independent Random Variables: General Multiparameter Case," *Journal of Multivariate Analysis*, 23, 183-208.

[7] Fu, Y. and R.N. Curnow (1990), "Maximum Likelihood Estimation of Multiple Change Points," *Biometrika*, 77, 563-73.

[8] Hamori, S. and A. Tokihisa (1997), "Testing for a Unit Root in the Presence of a Variance Shift," *Economics Letters*, 57, 245-253.

[9] Hsu, S. (1977), "Tests for Variance Shift at an Unknown Time Point," *Applied Statistics*, 26, 279-284.

[10] Inclán, C. (1993), "Detection of Multiple Changes of Variance Using Posterior Odds," *Journal of Business & Economic Statistics,* 11, 289-300.

[11] Leybourne, S.J., T.C. Mills and P. Newbold (1998) "Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break Under the Null," *Journal of Econometrics*, 87, 191-203.

[12] Nunes, L.C., C.M. Kuan and P. Newbold (1995), "Spurious Break," *Econometric Theory*, 11, 736-749.

[13] Picard, D. (1985), "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability*, 176, 841-867.

[14] Wichern, D.W., R. Miller and D. Hsu (1976), "Changes of Variance in First-Order Autoregressive Time Series Models-With an Application," *Applied Statistics*, 25, 248-256.

6 Appendix: Proofs of Theorems and Lemmas

6.1 Proof of Theorem 1.

First note that y_t can be written in the form

$$y_t = \sigma_{10} w_t 1[t \leq \tau_0 T] + (\sigma_{20} w_t - \lambda w_{\tau_0 T}) 1[t > \tau_0 T]$$

where $\lambda = \sigma_{20} - \sigma_{10}$ and $w_t = \sum_{i=1}^{t} \eta_i$. Also, t_c in (4) can be written as

$$t_c = \frac{T^{-1} \sum_{t=2}^{T} (y_{t-1} - \bar{y}) \varepsilon_t}{\hat{\sigma} \{T^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^2\}^{1/2}}.$$

Dealing with the numerator term first, we have

$$T^{-1} \sum_{2}^{T} (y_{t-1} - \bar{y}) \varepsilon_t = T^{-1} \sum_{2}^{T} y_{t-1} \varepsilon_t - T^{-1/2} \bar{y} T^{-1/2} \sum_{2}^{T} \varepsilon_t$$

and using the above representation for y_t, we find that

$$\begin{aligned} T^{-1} \sum_{2}^{T} y_{t-1} \varepsilon_t &= \sigma_{10}^2 T^{-1} \sum_{2}^{T} w_{t-1} \eta_t + (\sigma_{20}^2 - \sigma_{10}^2) T^{-1} \sum_{\tau_0 T+2}^{T} w_{t-1} \eta_t \\ &\quad - \sigma_{20} \lambda T^{-1/2} w_{\tau_0 T} T^{-1/2} (w_T - w_{\tau_0 T+1}) + o_p(1) \end{aligned}$$

where

$$\begin{aligned} T^{-1} \sum_{2}^{T} w_{t-1} \eta_t &\Rightarrow \frac{1}{2} \{W(1)^2 - 1\}, \\ T^{-1} \sum_{\tau_0 T+2}^{T} w_{t-1} \eta_t &\Rightarrow \frac{1}{2} \{W(1)^2 - W(\tau_0)^2 - (1 - \tau_0)\}, \\ T^{-1/2} w_{\tau_0 T} T^{-1/2} (w_T - w_{\tau_0 T+1}) &\Rightarrow W(\tau_0) \{W(1) - W(\tau_0)\}. \end{aligned}$$

Next,

$$T^{-1/2} \bar{y} = \sigma_{10} T^{-3/2} \sum_{1}^{T} w_t + \lambda T^{-3/2} \sum_{\tau_0 T+1}^{T} w_t - \lambda (1 - \tau_0) T^{-1/2} w_{\tau_0 T}$$

where

$$\begin{aligned} T^{-3/2} \sum_{1}^{T} w_t &\Rightarrow \int_0^1 W(r) dr, \\ T^{-3/2} \sum_{\tau_0 T+1}^{T} w_t &\Rightarrow \int_{\tau_0}^1 W(r) dr, \\ T^{-1/2} w_{\tau_0 T} &\Rightarrow W(\tau_0) \end{aligned}$$

and

$$T^{-1/2}\sum_{2}^{T}\varepsilon_t = \sigma_{10}T^{-1/2}w_T + \lambda T^{-1/2}(w_T - w_{\tau_0 T})$$

where

$$\begin{aligned} T^{-1/2}w_T &\Rightarrow W(1), \\ T^{-1/2}(w_T - w_{\tau_0 T}) &\Rightarrow W(1) - W(\tau_0). \end{aligned}$$

Gathering together these results then shows that

$$T^{-1}\sum_{2}^{T}(y_{t-1} - \bar{y})\varepsilon_t \Rightarrow \sigma_{10}^2 A(\delta_0, \tau_0) \tag{17}$$

Using a similar argument, we have

$$T^{-2}\sum_{2}^{T}(y_{t-1} - \bar{y})^2 \Rightarrow \sigma_{10}^2 B(\delta_0, \tau_0). \tag{18}$$

Finally, it is straightforward to show that

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 + o_p(1)$$

and

$$\begin{aligned} T^{-1}\sum_{1}^{T}\varepsilon_t^2 &= \sigma_{10}^2 T^{-1}\sum_{1}^{\tau_0 T+2}\eta_t^2 + \sigma_{20}^2 T^{-1}\sum_{\tau_0 T+1}^{T}\eta_t^2 \\ &= \sigma_{10}^2 T^{-1}\sum_{1}^{T}\eta_t^2 + (\sigma_{20}^2 - \sigma_{10}^2)(1-\tau_0)\{(1-\tau_0)T\}^{-1}\sum_{\tau_0 T+1}^{T}\eta_t^2 \\ &\xrightarrow{p} \sigma_{10}^2 + (\sigma_{20}^2 - \sigma_{10}^2)(1-\tau_0) \\ &= \tau_0\sigma_{10}^2 + (1-\tau_0)\sigma_{20}^2. \end{aligned}$$

So

$$\hat{\sigma}^2 \xrightarrow{p} \tau_0\sigma_{10}^2 + (1-\tau_0)\sigma_{20}^2 \tag{19}$$

Combining (17), (18) and (19) gives the result in (5).

6.2 Proof of Lemma 1.

First, the $T-$consistency $T(\hat{\tau}-\tau_0) = O_p(\lambda_0^{-2})$ is a direct result of Proposition 3 in Bai (1993). Without loss of generality, we consider the case $\hat{\tau} \leq \tau_0$. Then, $\hat{\sigma}_1^2 = (\hat{\tau}T)^{-1}\sum_1^{\hat{\tau}T}\varepsilon_t^2$ can be expressed as

$$\sigma_{10}^2 \frac{\tau_0}{\hat{\tau}}\{(\tau_0 T)^{-1}\sum_1^{\tau_0 T}\eta_t^2 - (\tau_0 T)^{-1}\sum_{\hat{\tau}T+1}^{\tau_0 T}\eta_t^2\}$$

where

$$\frac{\tau_0}{\hat{\tau}} \xrightarrow{p} 1,$$

$$(\tau_0 T)^{-1}\sum_1^{\tau_0 T}\eta_t^2 \xrightarrow{p} 1,$$

$$(\tau_0 T)^{-1}\sum_{\hat{\tau}T+1}^{\tau_0 T}\eta_t^2 \xrightarrow{p} 0.$$

This last result is obtained using the $T-$consistency of the break estimator: $T(\hat{\tau}-\tau_0) = O_p(\lambda_0^{-2})$. Hence, we have

$$\hat{\sigma}_1^2 - \sigma_{10}^2 = o_p(1).$$

The same argument can be applied to show $\hat{\sigma}_2^2 - \sigma_{20}^2 = o_p(1)$.

6.3 Proof of Lemma 2.

Consider

$$\begin{aligned}\left|V(\tau)-\hat{V}(\tau)\right| \leq& \left|(\tau T)^{-1}\sum_1^{\tau T}\log(\epsilon_t^2) - (\tau T)^{-1}\sum_{t=1}^{\tau T}\log(e_t^2)\right| \\ &+\left|\{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T}\log(\epsilon_t^2) - \{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T}\log(e_t^2)\right|.\end{aligned}$$

First we show the first term is $o_p(1)$. We define

$$\begin{aligned}f(\hat{\rho}) &= (\tau T)^{-1}\sum_{t=1}^{\tau T}\log(e_t^2) \\ &= (\tau T)^{-1}\sum_{t=1}^{\tau T}\log\{\epsilon_t^2 + a_t(T)\}\end{aligned}$$

where $a_t(T) = -2(\hat{\rho}-1)(y_{t-1}-\bar{y})\epsilon_t + (\hat{\rho}-1)^2(y_{t-1}-\bar{y})^2$. Then, by the mean value theorem, we have

$$f(\hat{\rho}) = f(1) + f^{'}(\tilde{\rho})(\hat{\rho}-1)$$

where $\tilde{\rho}$ is between $\hat{\rho}$ and 1. Hence, we have

$$\left|(\tau T)^{-1}\sum_{1}^{\tau T}\log(\epsilon_t^2)-(\tau T)^{-1}\sum_{t=1}^{\tau T}\log(e_t^2)\right|\le\left|(\tau T)^{-1}\sum_{1}^{\tau T}\frac{a_t(T)}{\tilde{e}_t^2(T)}\right|+|o_p(1)|.$$

where $\tilde{e}_t^2(T)=\epsilon_t^2-2(\tilde{\rho}-1)(y_{t-1}-\bar{y})\epsilon_t+(\tilde{\rho}-1)^2(y_{t-1}-\bar{y})^2$. Let $V_T=(\tau T)^{-1}\sum_{t=1}^{\tau T}\frac{a_t(T)}{\tilde{e}_t^2(T)}$. Let $\epsilon>0$ and $A=\{|V_T|>\epsilon\}$. Then, $P(|V_T|>\epsilon)=P(A)\le P(A\cap B)+P(B^c)$ for any event B. Choose $B=\{|\tilde{e}_t^2(T)|\ge c_T^2\}$ where (i) $c_T\to 0$ and (ii) $Tc_T^2\to\infty$. Then, $P(B^c)=P(|\tilde{e}_t^2(T)|<c_T^2)\to 0$ as $T\to\infty$. Hence, it is sufficient to show that $P(A\cap B)\to 0$. Consider

$$\begin{aligned}P(A\cap B) &= P(\left|(\tau T)^{-1}\sum_{t=1}^{T}\frac{a_t(T)}{\tilde{e}_t^2(T)}\right|>\epsilon \text{ and } \tilde{e}_t^2(T)\ge c_T^2)\\ &\le P((\tau Tc_T^2)^{-1}\left|\sum_{t=1}^{T}a_t(T)\right|>\epsilon).\end{aligned}$$

Since $(Tc_T^2)^{-1}=o(1)$ by the assumption (ii) on c_T, it is sufficient to show that $\left|\sum_{t=1}^{\tau T}a_t(T)\right|=O_p(1)$. Note that

$$\begin{aligned}\left|\sum_{t=1}^{\tau T}a_t(T)\right| &\le 2|\hat{\rho}-1|\left|\sum_{t=1}^{\tau T}(y_{t-1}-\bar{y})\epsilon_t\right|+(\hat{\rho}-1)^2\left|\sum_{t=1}^{\tau T}(y_{t-1}-\bar{y})^2\right|+o_p(1)\\ &= 2|T(\hat{\rho}-1)|\left|(\tau T)^{-1}\sum_{t=1}^{\tau T}(y_{t-1}-\bar{y})\epsilon_t\right|\tau\\ &\quad +[T(\hat{\rho}-1)]^2\left|(\tau T)^{-2}\sum_{t=1}^{\tau T}(y_{t-1}-\bar{y})^2\right|\tau^2+o_p(1)\\ &= O_p(1)+O_p(1)+o_p(1)=O_p(1).\end{aligned}$$

In a very similar way it can be shown that $|\{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T}\log(\epsilon_t^2)-\{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T}\log(e_t^2)|=o_p(1)$. Therefore, $\left|V(\tau)-\hat{V}(\tau)\right|=o_p(1)$.

6.4 Proof of Lemma 4.

First, note that t_{cM} can be expressed as

$$\begin{aligned}t_{cM} &= \frac{(\tau_0 T)^{-1}\sum_2^{\tau_0 T}\breve{y}_{t-1}\triangle\breve{y}_t+\{(1-\tau_0)T\}^{-1}\sum_{\tau_0 T+1}^{T}\breve{y}_{t-1}\triangle\breve{y}_t}{[(\tau_0 T)^{-2}\sum_2^{\tau_0 T}\breve{y}_{t-1}^2+\{(1-\tau_0)T\}^{-2}\sum_{\tau_0 T+1}^{T}\breve{y}_{t-1}^2]^{1/2}}\\ &= \frac{(\tau_0 T)^{-1}\sum_2^{\tau_0 T}(w_{t-1}-\bar{w}_{(1)})\triangle w_t+\{(1-\tau_0)T\}^{-1}\sum_{\tau_0 T+2}^{T}(w_{t-1}-\bar{w}_{(2)})\triangle w_t}{\left[(\tau_0 T)^{-2}\sum_2^{\tau_0 T}(w_{t-1}-\bar{w}_{(1)})^2+\{(1-\tau_0)T\}^{-2}\sum_{\tau_0 T+2}^{T}(w_{t-1}-\bar{w}_{(2)})^2\right]^{1/2}}+o_p(1).\end{aligned}$$

Then, it is straightforward to show that

$$\begin{aligned}
(\tau_0 T)^{-1}\sum_{2}^{\tau_0 T}(w_{t-1}-\bar{w}_{(1)})\triangle w_t &\Rightarrow G_1,\\
\{(1-\tau_0)T\}^{-1}\sum_{\tau_0 T+2}^{T}(w_{t-1}-\bar{w}_{(2)})\triangle w_t &\Rightarrow G_2,\\
(\tau_0 T)^{-2}\sum_{2}^{\tau_0 T}(w_{t-1}-\bar{w}_{(1)})^2 &\Rightarrow H_1,\\
\{(1-\tau_0)T\}^{-2}\sum_{\tau_0 T+2}^{T}(w_{t-1}-\bar{w}_{(2)})^2 &\Rightarrow H_2
\end{aligned}$$

which delivers the desired result.

6.5 Proof of Theorem 2.

We write t_{cF} as

$$t_{cF}=\frac{(\hat{\tau}T)^{-1}\sum_2^{\hat{\tau}T}\breve{y}_{t-1}\triangle\breve{y}_t+\{(1-\hat{\tau})T\}^{-1}\sum_{\hat{\tau}T+1}^{T}\breve{y}_{t-1}\triangle\breve{y}_t}{\left[(\hat{\tau}T)^{-2}\sum_2^{\hat{\tau}T}\breve{y}_{t-1}^2+\{(1-\hat{\tau})T\}^{-2}\sum_{\hat{\tau}T+1}^{T}\breve{y}_{t-1}^2\right]^{1/2}}.$$

We will only prove here that $(\hat{\tau}T)^{-1}\sum_1^{\hat{\tau}T}\breve{y}_{t-1}\triangle\breve{y}_t\Rightarrow G_1$. The same arguments can be applied to the other terms in t_{cF}. Note that

$$(\hat{\tau}T)^{-1}\sum_{2}^{\hat{\tau}T}\breve{y}_{t-1}\triangle\breve{y}_t=(\hat{\tau}T)^{-1}\sum_{2}^{\hat{\tau}T}\tilde{y}_{t-1}\triangle\tilde{y}_t-(\hat{\tau}T)^{-3/2}\sum_{2}^{\hat{\tau}T}\tilde{y}_t(\hat{\tau}T)^{-1/2}\sum_{2}^{\hat{\tau}T}\triangle\tilde{y}_t$$

Without loss of generality, we consider the case $\hat{\tau}\le\tau_0$. Then, the first term can be expressed as

$$(\hat{\tau}T)^{-1}\sum_{2}^{\hat{\tau}T}\tilde{y}_{t-1}\triangle\tilde{y}_t=\frac{\tau_0}{\hat{\tau}}\frac{\sigma_{10}^2}{\hat{\sigma}_1^2}(\tau_0T)^{-1}\sum_{2}^{\tau_0T}w_{t-1}\triangle w_t-\frac{\tau_0}{\hat{\tau}}\frac{\sigma_{10}^2}{\hat{\sigma}_1^2}(\tau_0T)^{-1}\sum_{\hat{\tau}T+1}^{\tau_0T}w_{t-1}\triangle w_t$$

where

$$\frac{\tau_0}{\hat{\tau}}\xrightarrow{p}1,$$

$$\frac{\sigma_{10}^2}{\hat{\sigma}_1^2}\xrightarrow{p}1,$$

$$(\tau_0T)^{-1}\sum_{2}^{\tau_0T}w_{t-1}\triangle w_t\Rightarrow(1/2)\{W_1(1)^2-1\},$$

$$(\tau_0T)^{-1}\sum_{\hat{\tau}T+1}^{\tau_0T}w_{t-1}\triangle w_t\xrightarrow{p}0.$$

The last result is obtained using the $T-$consistency of the break estimator: $T(\hat{\tau}-\tau_0)=O_p(\lambda_0^{-2})$. Hence, we have

$$(\hat{\tau}T)^{-1}\sum_{2}^{\hat{\tau}T}\tilde{y}_{t-1}\triangle\tilde{y}_t \Rightarrow (1/2)\{W_1(1)^2-1\}.$$

The limit of the other two terms are given by

$$\begin{aligned}
(\hat{\tau}T)^{-3/2}\sum_{2}^{\hat{\tau}T}\tilde{y}_t &= (\frac{\tau_0}{\hat{\tau}})^{3/2}\frac{\sigma_{10}}{\hat{\sigma}_1}(\tau_0T)^{-3/2}\sum_{2}^{\tau_0T}w_t-(\frac{\tau_0}{\hat{\tau}})^{3/2}\frac{\sigma_{10}}{\hat{\sigma}_1}(\tau_0T)^{-3/2}\sum_{\hat{\tau}T+1}^{\tau_0T}w_t\\
&\Rightarrow \int_0^1 W_1(r),\\
(\hat{\tau}T)^{-1/2}\sum_{2}^{\hat{\tau}T}\triangle\tilde{y}_t &= (\frac{\tau_0}{\hat{\tau}})^{1/2}\frac{\sigma_{10}}{\hat{\sigma}_1}(\tau_0T)^{-1/2}\sum_{2}^{\tau_0T}\eta_t-(\frac{\tau_0}{\hat{\tau}})^{1/2}\frac{\sigma_{10}}{\hat{\sigma}_1}(\tau_0T)^{-1/2}\sum_{\hat{\tau}T+1}^{\tau_0T}\eta_t\\
&\Rightarrow W_1(1)
\end{aligned}$$

using the same argument. Therefore,

$$(\hat{\tau}T)^{-1}\sum_{2}^{\hat{\tau}T}\breve{y}_{t-1}\triangle\breve{y}_t \Rightarrow G_1$$

A similar approach establishes that

$$\begin{aligned}
\{(1-\hat{\tau})T\}^{-1}\sum_{\hat{\tau}T+1}^{T}\breve{y}_{t-1}\triangle\breve{y}_t &\Rightarrow G_2,\\
(\hat{\tau}T)^{-2}\sum_{1}^{\hat{\tau}T}\breve{y}_{t-1}^2 &\Rightarrow H_1,\\
\{(1-\hat{\tau})T\}^{-2}\sum_{\hat{\tau}T+1}^{T}\breve{y}_{t-1}^2 &\Rightarrow H_2
\end{aligned}$$

which completes the proof.

6.6 Proof of Theorem 3.

Consider $\tau\in[\tau_1,\tau_2]$ and define

$$\begin{aligned}
\hat{\sigma}_1^2(\tau) &= (\tau T)^{-1}\sum_{1}^{\tau T}e_t^2,\\
\hat{\sigma}_2^2(\tau) &= \{(1-\tau)T\}^{-1}\sum_{\tau T+1}^{T}e_t^2.
\end{aligned}$$

Then, it is straightforward to show that $\hat{\sigma}_1^2(\tau) \xrightarrow{p} \sigma_0^2$ and $\hat{\sigma}_2^2(\tau) \xrightarrow{p} \sigma_0^2$ uniformly in τ. Define

$$\breve{y}_t = (\tilde{y}_t - \bar{y}_{(1)})1[t \leq \tau_0 T] + (\tilde{y}_t - \bar{y}_{(2)})1[t > \tau_0 T]$$

where

$$\tilde{y}_t = \hat{\sigma}_1^2(\tau)^{-1} y_t 1[t \leq \tau T] + \hat{\sigma}_2^2(\tau)^{-1} y_t 1[t > \tau T],$$

$$\bar{y}_{(1)} = (\tau T)^{-1} \sum_1^{\tau T} \tilde{y}_t,$$

$$\bar{y}_{(2)} = \{(1-\tau)T\}^{-1} \sum_{\tau T+1}^{T} \tilde{y}_t.$$

Then, the FMGLS statistic based on τ is given by

$$t_{cF}(\tau) = \frac{(\tau T)^{-1} \sum_2^{\tau T} \breve{y}_{t-1} \triangle \breve{y}_t + \{(1-\tau)T\}^{-1} \sum_{\tau T+1}^{T} \breve{y}_{t-1} \triangle \breve{y}_t}{[(\tau T)^{-2} \sum_2^{\tau T} \breve{y}_{t-1}^2 + \{(1-\tau)T\}^{-2} \sum_{\tau T+1}^{T} \breve{y}_{t-1}^2]^{1/2}}.$$

Using the Functional Central Limit Theorem (FCLT) it can be easily shown that

$$t_{cF}(\tau) \Rightarrow \Upsilon = \frac{G_1 + G_2}{(H_1 + H_2)^{1/2}} \tag{20}$$

uniformly in τ. By Nunes, Kuan and Newbold (1995) and Bai (1998), the break estimator $\hat{\tau}$ obtained by searching $[\tau_1, \tau_2]$ converges in distribution to a well-behaved limiting random variable

$$\hat{\tau} \Rightarrow \hat{\tau}_\infty \tag{21}$$

where $\hat{\tau}_\infty$ is a solution to some quadratic maximization problem. Note that $t_{cF}(\hat{\tau}) \Rightarrow \Upsilon$ if and only if for every bounded continuous function φ, $E[\varphi\{t_{cF}(\hat{\tau})\}] \to E\{\varphi(\Upsilon)\}$. Let φ be a bounded and continuous function and ϵ be a positive real number. We need to find an $N(\epsilon) > 0$ such that $|E[\varphi\{t_{cF}(\hat{\tau})\}] - E\{\varphi(\Upsilon)\}| < \varepsilon$ for all $T > N(\epsilon)$. First, note that *(i)* there exists an $N_1(\epsilon) > 0$ such that $\sup_\tau |E[\varphi\{t_{cF}(\tau)\}] - E\{\varphi(\Upsilon)\}| < \varepsilon$ for all $T > N_1(\epsilon)$ by (20) and *(ii)* there exits an $N_2 > 0$ such that $\hat{\tau}$ has a proper density, denoted by $f_{\hat{\tau}}(\tau)$, for all $T > N_2$ by (21). Consider $T > N_2$:

$$\begin{aligned} |E[\varphi\{t_{cF}(\hat{\tau})\}] - E\{\varphi(\Upsilon)\}| &= |\int_{\tau_1}^{\tau_2} E[\{\varphi\{t_{cF}(\hat{\tau})\} - \varphi(\Upsilon)\}|_{\hat{\tau}=\tau}] f_{\hat{\tau}}(\tau) d\tau| \\ &\leq \int_{\tau_1}^{\tau_2} \sup_\tau |E[\{\varphi\{t_{cF}(\hat{\tau})\} - \varphi(\Upsilon)\}|_{\hat{\tau}=\tau}]| f_{\hat{\tau}}(\tau) d\tau \\ &\leq \int_{\tau_1}^{\tau_2} \epsilon f_{\hat{\tau}}(\tau) d\tau \\ &= \epsilon \end{aligned}$$

for all $T > N_1(\epsilon)$. Therefore, we choose $N(\epsilon) = \max\{N_1(\epsilon), N_2\}$ which completes the proof.

Table 1. Critical Values of t_{cF}.

T	10%	5%	1%
100	-2.98	-3.26	-3.82
200	-3.00	-3.28	-3.83
400	-3.03	-3.32	-3.85
∞	-3.04	-3.33	-3.86

Table 2. Size of t_{cF} at Asymptotic 5%-level Critical Values, $T = 100$.

τ_0 \ δ_0	1.00	0.80	0.60	0.40	0.25
0.15	0.045	0.053	0.062	0.070	0.071
0.30	0.045	0.055	0.064	0.079	0.083
0.45	0.045	0.049	0.060	0.073	0.074
0.60	0.045	0.046	0.057	0.066	0.067
0.75	0.045	0.043	0.046	0.054	0.057

Table 3. Size of t_{cF} at Asymptotic 5%-level Critical Values, $T = 200$.

τ_0 \ δ_0	1.00	0.80	0.60	0.40	0.25
0.15	0.048	0.054	0.065	0.067	0.066
0.30	0.048	0.052	0.063	0.065	0.069
0.45	0.048	0.049	0.063	0.068	0.067
0.60	0.048	0.047	0.056	0.065	0.064
0.75	0.048	0.045	0.049	0.058	0.055

Table 4. Power of t_{cF} at Asymptotic 5%-level Critical Values, $T = 100$.

τ_0	δ_0	1.00	0.80	0.60	0.40	0.25
0.15	$\rho = 0.9$	0.181	0.185	0.185	0.228	0.343
	$\rho = 0.8$	0.501	0.504	0.502	0.516	0.586
0.30	$\rho = 0.9$	0.181	0.190	0.217	0.272	0.367
	$\rho = 0.8$	0.501	0.520	0.600	0.677	0.748
0.45	$\rho = 0.9$	0.181	0.193	0.233	0.286	0.378
	$\rho = 0.8$	0.501	0.544	0.653	0.757	0.835
0.60	$\rho = 0.9$	0.181	0.193	0.225	0.283	0.342
	$\rho = 0.8$	0.501	0.540	0.643	0.742	0.797
0.75	$\rho = 0.9$	0.181	0.191	0.208	0.225	0.240
	$\rho = 0.8$	0.501	0.522	0.556	0.603	0.638

Table 5. Power of t_{cF} at Asymptotic 5%-level Critical Values, $T = 200$.

τ_0	δ_0	1.00	0.80	0.60	0.40	0.25
0.15	$\rho = 0.9$	0.503	0.503	0.502	0.515	0.586
	$\rho = 0.8$	0.811	0.813	0.821	0.857	0.887
0.30	$\rho = 0.9$	0.503	0.541	0.616	0.692	0.789
	$\rho = 0.8$	0.811	0.876	0.956	0.996	0.999
0.45	$\rho = 0.9$	0.503	0.559	0.692	0.771	0.844
	$\rho = 0.8$	0.811	0.890	0.978	1.000	1.000
0.60	$\rho = 0.9$	0.503	0.556	0.673	0.746	0.802
	$\rho = 0.8$	0.811	0.886	0.972	0.996	1.000
0.75	$\rho = 0.9$	0.503	0.527	0.564	0.608	0.650
	$\rho = 0.8$	0.811	0.865	0.919	0.966	0.988

Table 6. Size of t_{cF} at Asymptotic 5%-level Critical Values, $T = 100$, ARIMA(1,1,0) DGP.

	δ_0	1.00	0.80	0.60	0.40	0.25
τ_0						
0.15		0.052	0.063	0.072	0.078	0.072
0.30		0.052	0.063	0.075	0.085	0.079
0.45		0.052	0.062	0.074	0.081	0.074
0.60		0.052	0.058	0.070	0.072	0.070
0.75		0.052	0.051	0.061	0.068	0.062

Table 7. Size of t_{cF} at Asymptotic 5%-level Critical Values, $T = 200$, ARIMA(1,1,0) DGP.

	δ_0	1.00	0.80	0.60	0.40	0.25
τ_0						
0.15		0.051	0.060	0.071	0.069	0.071
0.30		0.051	0.063	0.072	0.073	0.071
0.45		0.051	0.059	0.066	0.069	0.067
0.60		0.051	0.056	0.063	0.068	0.066
0.75		0.051	0.056	0.059	0.065	0.056

Table 8. Critical Values of t_{tF}.

	10%	5%	1%
T			
100	−3.71	−3.98	−4.51
200	−3.79	−4.06	−4.57
400	−3.84	−4.11	−4.63
∞	−3.86	−4.13	−4.65

Figure 1. Approximate Expected Value of t_c with Decrease in Variance.

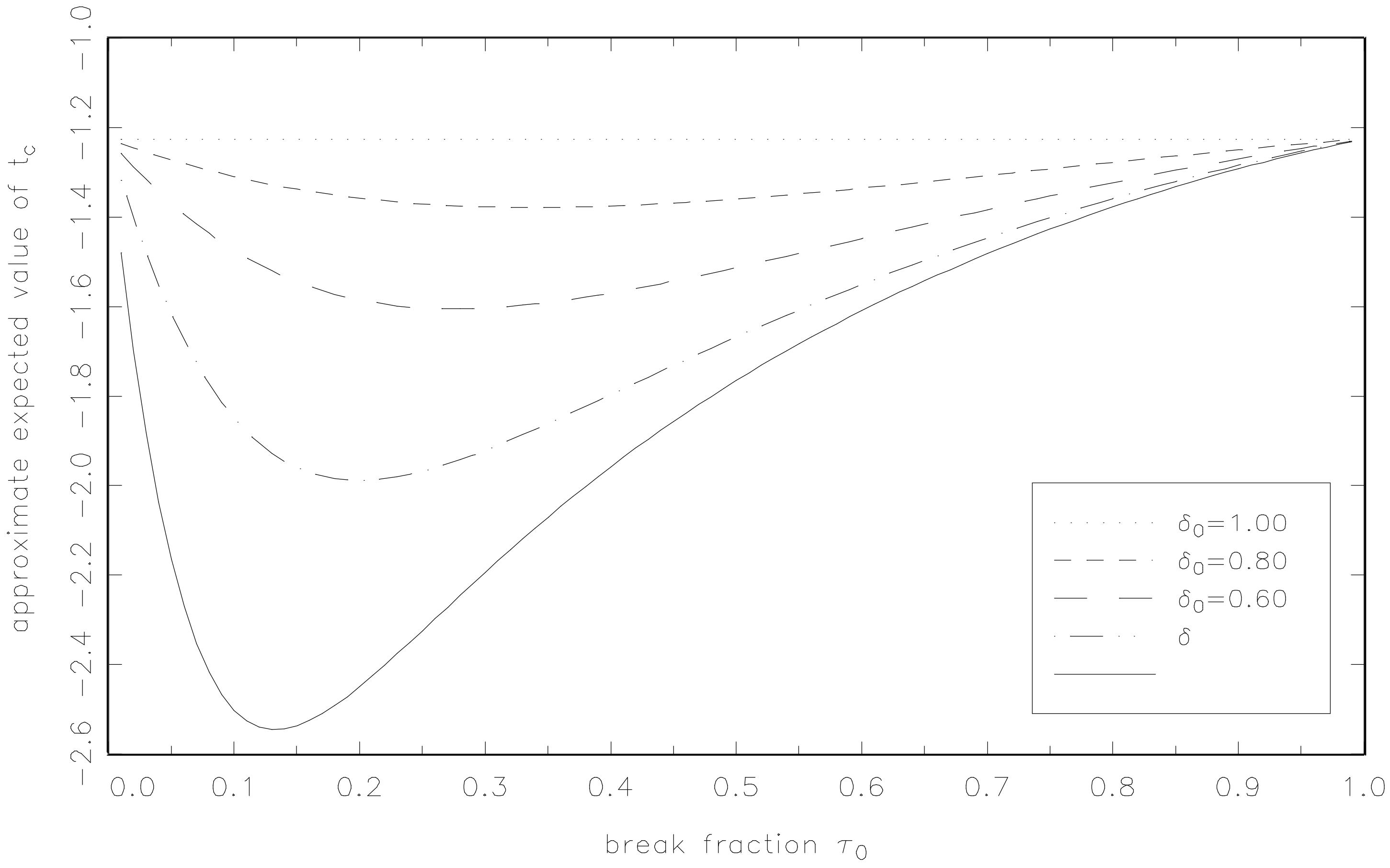

Figure 2. Actual Size of t_c with Decrease in Variance, T = 100.

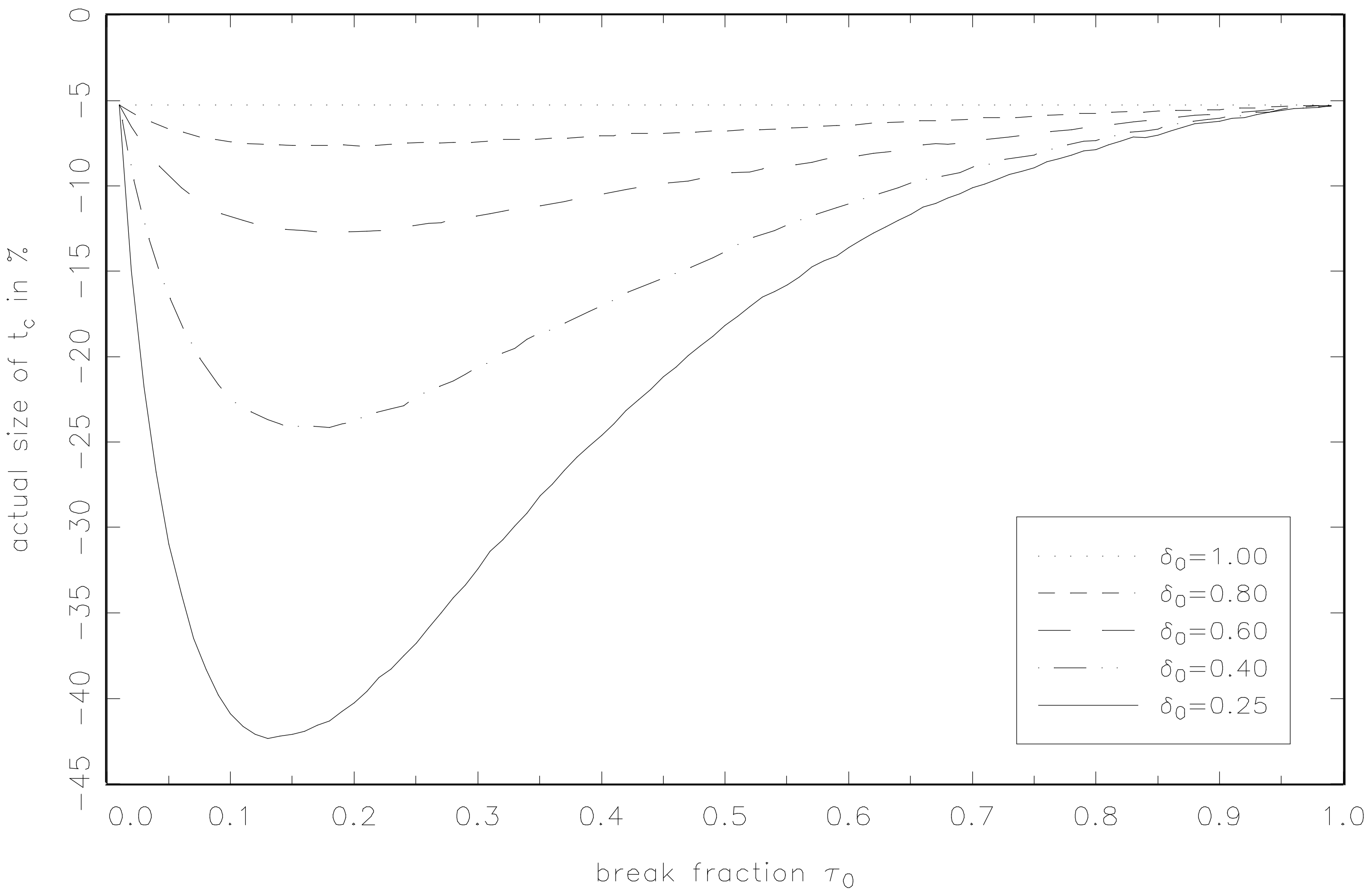

Figure 3. Actual Size of t_t with Decrease in Variance, T = 100.