UvA ECONOMETRICS

Discussion Paper: 2008/03

Nuisance parameter free inference on cointegration parameters in the presence of a variance shift

H. Peter Boswijk

www.feb.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Nuisance parameter free inference on cointegration parameters in the presence of a variance shift

H. Peter Boswijk*
Department of Quantitative Economics, University of Amsterdam,
Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands

November 30, 2008

Abstract

Kourogenis and Pittis (2008) show that the presence of a variance shift implies that the OLS t-statistic in a triangular cointegrated model displays asymptotic size distortions. For the same model, this paper provides two simple solutions to the size problems, the first based on White (1980) standard errors, the second based on the wild bootstrap.

Keywords: Cointegration; Heteroskedasticity; Variance shifts; Wild bootstrap

JEL classification: C32

1 Introduction

Recent research has indicated that unit root and cointegration tests are substantially affected by persistent changes in innovation variances; see Cavaliere and Taylor (2007, 2008), Cavaliere *et al.* (2009), and the references therein. This is relevant for macro-economic time series, many of which display a decreasing volatility during the 1980s (known as the *Great Moderation*); and for financial time series, which often display slow volatility mean reversion. In a recent paper, Kourogenis and Pittis (2008) show that inference on cointegration parameters is also affected by variance shifts. In particular, they show that in a triangular cointegrated system with no serial dependence or endogeneity, a variance shift leads to size distortions in the OLS t-statistic, both asymptotically and in finite samples.

This paper investigates the possibility to correct these size distortions, in the same model as analysed by Kourogenis and Pittis (2008), following two approaches. First, we show that using White (1980)'s heteroskedasticity-consistent standard errors yields asymptotically standard normal inference. Secondly, inspired by Cavaliere and Taylor (2008) and Cavaliere *et al.* (2009), we show that the wild bootstrap provides an equally effective approach to solve the size problems. We prove the asymptotic validity of both approaches, and compare their effectiveness in a small Monte Carlo experiment.

*Corresponding author. Tel: +31 20 5254316; fax: +31 20 5254349; e-mail: h.p.boswijk@uva.nl

Throughout the paper, we use the following notation. Weak convergence (convergence in distribution) and convergence in probability are denoted by $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{P}$, respectively, whereas $\xrightarrow{\mathcal{L}}_P$ denotes weak convergence in probability, see Giné and Zinn (1990). Equality in distribution is denoted by $\equiv$. The indicator function of the set A is denoted $\mathbf{1}_A(\cdot)$, and $\lfloor x \rfloor$ denotes the integer part of x.

2 The model and OLS inference

Consider the following triangular cointegrated system, analysed by Kourogenis and Pittis (2008) (henceforth KP08) for a bivariate time series $\{(y_t, x_t)'\}_{t=1}^T$:

$$y_t = \theta x_t + u_{1t}, \tag{1}$$

$$\Delta x_t = u_{2t}, \tag{2}$$

where $\{u_t = (u_{1t}, u_{2t})'\}_{t=1}^T$ is a sequence satisfying $u_t = \Sigma_t^{1/2} z_t$, where $z_t \sim$ i.i.d. $(0, I_2)$, and

$$\Sigma_t = \mathrm{var}(u_t) = \left\{ \mathbf{1}_{[0,s)}\left(\tfrac{t}{T}\right) + a\mathbf{1}_{[s,1]}\left(\tfrac{t}{T}\right) \right\} \Sigma, \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}, \tag{3}$$

with a is a positive scalar. The specification (3) implies a proportional shift in the variance matrix of u_t at time $t = \lfloor sT \rfloor$: it changes from $\mathrm{var}(u_t) = \Sigma$ for $1 \le t < \lfloor sT \rfloor$, to $\mathrm{var}(u_t) = a\Sigma$ for $\lfloor sT \rfloor \le t \le T$.

Define the least-squares estimator $\hat{\theta}$ in (1), and the corresponding t-statistic for testing $H_0 : \theta = \theta_0$:

$$\hat{\theta} = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2}, \qquad t_{ols} = \frac{\sqrt{\sum_{t=1}^T x_t^2}(\hat{\theta} - \theta_0)}{\hat{\sigma}_1}, \tag{4}$$

where $\hat{\sigma}_1^2 = T^{-1} \sum_{t=1}^T \hat{u}_{1t}^2$, with $\hat{u}_{1t} = y_t - \hat{\theta} x_t$.

Theorem 1 *In the model (1)–(3), under H_0 and as $T \to \infty$,*

$$T(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} \frac{\sigma_1}{\sigma_2} \frac{\int_0^1 X(r)Q(r)dW_1(r)}{\int_0^1 X(r)^2 dr}, \tag{5}$$

$$t_{ols} \xrightarrow{\mathcal{L}} \frac{\int_0^1 X(r)Q(r)dW_1(r)}{\sqrt{\int_0^1 Q(r)^2 dr \int_0^1 X(r)^2 dr}} \equiv \sqrt{A} Z, \tag{6}$$

where $(W_1(\cdot), W_2(\cdot))'$ is a bivariate standard Brownian motion, $Q(r) = \mathbf{1}_{[0,s)}(r) + \sqrt{a}\mathbf{1}_{[s,1]}(r)$, $X(r) = \int_0^r Q(u)dW_2(u)$,

$$A = \frac{\int_0^1 X(r)^2 Q(r)^2 dr}{\int_0^1 Q(r)^2 dr \int_0^1 X(r)^2 dr}, \tag{7}$$

and $Z \sim N(0,1)$, independent of A.

Proof. See KP08, Theorems 1 and 2, except for an error in these theorems, which is corrected here. Define

$$W_T(r) = \begin{pmatrix} W_{1T}(r) \\ W_{2T}(r) \end{pmatrix} = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \Sigma_t^{-1/2} u_t = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} z_t.$$

Because $z_t = (z_{1t}, z_{2t})' \sim$ i.i.d. $(0, I_2)$, the invariance principle implies that $W_T(\cdot) \xrightarrow{\mathcal{L}} W(\cdot) = (W_1(\cdot), W_2(\cdot))'$. Define $X_T(r) = (\sigma_2^2 T)^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} u_{2t}$. Because $u_{2t} = \sigma_2 Q\left(\frac{t}{T}\right) z_{2t}$, we have

$$X_T(r) = \int_0^r Q(u) dW_{2T}(u) \xrightarrow{\mathcal{L}} \int_0^r Q(u) dW_2(u) = X(r), \qquad r \in [0,1],$$

which follows from the continuous mapping theorem, because $Q(r)$ is of bounded variation. Note that $X(\cdot)$ is a Gaussian process with mean zero and variance

$$\mathrm{var}(X(r)) = \int_0^r Q(u)^2 du = r\mathbf{1}_{[0,s)}(r) + (s + a(r-s))\mathbf{1}_{[s,1]}(r). \tag{8}$$

Therefore, defining

$$D(r) = \sqrt{\frac{\mathrm{var}(X(r))}{r}} = \mathbf{1}_{[0,s)}(r) + \sqrt{a + \frac{s(1-a)}{r}}\mathbf{1}_{[s,1]}(r),$$

and $U(r) = X(r)/D(r)$, it follows that $U(r) \sim N(0, r)$. KP08 relabel the process $U(\cdot)$ as $W_2(\cdot)$, and claim that it is a standard Brownian motion. However, although $U(r)$ has the distribution of a Brownian motion at any time $r \in [0,1]$, it does not have independent increments. In particular, for any $u \in [s,1]$, we have

$$U(u) - U(s) = \frac{W_2(s) + \sqrt{a}[W_2(u) - W_2(s)]}{\sqrt{a + s(1-a)/r}} - W_2(s),$$

which is clearly not independent of $U(s) - U(0) = W_2(s)$. Replacing $D(r)W_2(r)$ in the expressions of KP08 by $X(r)$, we obtain

$$\frac{1}{T}\sum_{t=1}^{T} x_t u_{1t} \quad \xrightarrow{\mathcal{L}} \quad \sigma_1\sigma_2 \int_0^1 X(r)Q(r)dW_1(r), \tag{9}$$

$$\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 \quad \xrightarrow{\mathcal{L}} \quad \sigma_2^2 \int_0^1 X(r)^2 dr, \tag{10}$$

which leads to (5), and

$$\hat{\sigma}_1^2 \xrightarrow{P} \sigma_1^2 \int_0^1 Q(r)^2 dr = \sigma_1^2(a + s(1-a)), \tag{11}$$

which leads to (6) and (7). □

3 Nuisance parameter free inference

In this section two solutions are proposed to the nuisance parameter problem implied by Theorem 1. Both methods build on an auxiliary result, given in Lemma 1. Let $\sigma_{1t}^2 = \mathrm{var}(u_{1t}) = \sigma_1^2 Q\left(\frac{t}{T}\right)^2$, and define the integrated variance process

$$V(r) = \sigma_1^2 \int_0^r Q(u)^2 du = \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} \sigma_{1t}^2 + \left(r - \frac{\lfloor rT \rfloor}{T}\right)\sigma_{1,\lfloor rT \rfloor + 1}^2, \qquad r \in [0,1].$$

A natural estimator of $V(r)$ based on OLS residuals is given by

$$\hat{V}_T(r) = \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor} \hat{u}_{1t}^2 + \left(r - \frac{\lfloor rT \rfloor}{T}\right)\hat{u}_{1,\lfloor rT \rfloor + 1}^2, \qquad r \in [0,1].$$

Cavaliere and Taylor (2007) call $\eta(\cdot) = V(\cdot)/V(1)$ the *variance profile*, and prove uniform consistency of the estimator $\hat{\eta}_T(\cdot) = \hat{V}_T(\cdot)/\hat{V}_T(1)$ in a unit-root autoregression with residual autocorrelation. Lemma 1 provides the corresponding result for the present model.

Lemma 1 *In the model (1)–(3) with $\kappa_i = E(z_{it}^4) < \infty, i = 1, 2,$ as $T \to \infty$,*

$$\sup_{r\in[0,1]} \left|\hat{V}_T(r) - V(r)\right| \xrightarrow{P} 0.$$

Proof. Consider first pointwise consistency, for fixed r. Let $V_T(\cdot)$ be the (infeasible) version of $\hat{V}_T(\cdot)$ with $\{\hat{u}_{1t}\}$ replaced by $\{u_{1t}\}$, and let $h_T(r) = r - \lfloor rT \rfloor/T$. Then

$$V_T(r) - V(r) = \frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}(u_{1t}^2 - \sigma_{1t}^2) + h_T(r)(u_{1,\lfloor rT\rfloor+1}^2 - \sigma_{1,\lfloor rT\rfloor+1}^2).$$

Using $u_{1t}^2 - \sigma_{1t}^2 = \sigma_{1t}^2(z_t^2 - 1)$, it follows that $V_T(r) - V(r)$ has mean 0 and variance

$$\frac{1}{T^2}\sum_{t=1}^{\lfloor rT \rfloor}\sigma_{1t}^4(\kappa_1 - 1) + h_T(r)^2\sigma_{1,\lfloor rT\rfloor+1}^4(\kappa_1 - 1) \to 0,$$

because $h_T(r) < T^{-1}$ and σ_{1t}^2 is bounded. This implies $|V_T(r) - V(r)| \xrightarrow{P} 0$. Next, using $\hat{u}_{1t} = u_{1t} - (\hat{\theta} - \theta)x_t$, we have

$$\begin{aligned}\hat{V}_T(r) - V_T(r) &= -2(\hat{\theta} - \theta)\left(\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}x_t u_{1t} + h_T(r)x_{\lfloor rT\rfloor+1}u_{1,\lfloor rT\rfloor+1}\right) \\ &\quad +(\hat{\theta} - \theta)^2\left(\frac{1}{T}\sum_{t=1}^{\lfloor rT \rfloor}x_t^2 + h_T(r)x_{\lfloor rT\rfloor+1}^2\right).\end{aligned}$$

Both right-hand side terms converge in probability to zero because $\hat{\theta} - \theta = O_P(T^{-1})$, $\sum_{t=1}^{\lfloor rT \rfloor} x_t u_{1t} = O_P(T)$ and $\sum_{t=1}^{\lfloor rT \rfloor} x_t^2 = O_P(T^2)$, see (9) and (10). Therefore

$$\left|\hat{V}_T(r) - V(r)\right| \leq \left|\hat{V}_T(r) - V_T(r)\right| + |V_T(r) - V(r)| \xrightarrow{P} 0.$$

This is strengthened to uniform convergence in probability because both $\hat{V}_T(\cdot)$ and $V(\cdot)$ are continuous and monotonically increasing in r, and $V(\cdot)$ is bounded. $\square$

Based on this result, a first possible solution to correct for a shift in variance is to replace the OLS standard error of $\hat{\theta}$ by the White (1980) heteroskedasticity-consistent standard error s_W, given by

$$s_W^2 = \frac{\sum_{t=1}^T x_t^2\hat{u}_{1t}^2}{\left(\sum_{t=1}^T x_t^2\right)^2}.$$

The resulting t-statistic is $t_W = (\hat{\theta} - \theta_0)/s_W$.

Theorem 2 *Under the conditions of Lemma 1, under H_0 and as $T \to \infty$, $t_W \xrightarrow{\mathcal{L}} N(0, 1)$.*

Proof. It is convenient to define for $r \in [0, 1)$,

$$X_T^*(r) = \frac{1}{\sigma_2 \sqrt{T}} x_{\lfloor rT \rfloor + 1} = X_T(r) + \frac{x_0 + u_{\lfloor rT \rfloor + 1}}{\sigma_2 \sqrt{T}} = X_T(r) + o_P(1)$$

(we may take $X_T^*(1) = \lim_{r \uparrow 1} X_T^*(r)$). Then we find

$$\frac{1}{T^2} \sum_{t=1}^{T} x_t^2 \hat{u}_{1t}^2 = \sigma_2^2 \int_0^1 X_T^*(r)^2 d\hat{V}_T(r). \tag{12}$$

Because $(X_T^*(\cdot), \hat{V}_T(\cdot)) \xrightarrow{\mathcal{L}} (X(\cdot), V(\cdot))$, and $V(r)$ is of bounded variation, so that $\int_0^1 X(r)^2 dV(r)$ is defined as a Lebesgue-Stieltjes integral, the continuous mapping theorem implies

$$\frac{1}{T^2} \sum_{t=1}^{T} x_t^2 \hat{u}_{1t}^2 \xrightarrow{\mathcal{L}} \sigma_2^2 \int_0^1 X(r)^2 dV(r) = \sigma_1^2 \sigma_2^2 \int_0^1 X(r)^2 Q(r)^2 dr. \tag{13}$$

Therefore,

$$t_W = \frac{T^{-1} \sum_{t=1}^{T} x_t u_{1t}}{\sqrt{T^{-2} \sum_{t=1}^{T} x_t^2 \hat{u}_{1t}^2}} \xrightarrow{\mathcal{L}} \frac{\sigma_1 \sigma_2 \int_0^1 X(r) Q(r) dW_1(r)}{\sqrt{\sigma_1^2 \sigma_2^2 \int_0^1 X(r)^2 Q(r)^2 dr}},$$

which has a standard normal distribution because $W_2(\cdot)$ is independent of $X(\cdot)$. $\square$

A second approach to deliver standard normal inference is based on the wild bootstrap, see, e.g., Mammen (1993). This approach has recently been applied successfully to unit root and cointegration testing problems in the presence of nonstationary volatility by Cavaliere and Taylor (2008) and Cavaliere *et al.* (2009). The idea is to approximate the limiting null distribution of t_{ols} as given in Theorem 1 by the sampling distribution of $\{t_{ols}^b, b = 1, \ldots, B\}$, where for each b, t_{ols}^b is given by (4) with y_t replaced by $y_t^b = \theta_0 x_t + \hat{u}_{1t} w_t^b$, and $\{w_t^b\}_{t=1}^T$ is an i.i.d. $N(0,1)$ sequence, independent of $\{y_t, x_t\}_{t=1}^T$ and independent across b. The test rejects if the bootstrap p-value $\# \left\{ \left| t_{ols}^b \right| > |t_{ols}| \right\} / B$ is less than the nominal significance level. The asymptotic validity of this test is implied by Theorem 3.

Theorem 3 *Under the conditions of Lemma 1, under H_0 and as $T \to \infty$, $t_{ols}^b \xrightarrow{\mathcal{L}}_P \sqrt{A} Z$, where A is defined in (7) and $Z \sim N(0,1)$, independent of A.*

Proof. Write t_{ols}^b as

$$t_{ols}^b = \frac{\sum_{t=1}^{T} x_t \hat{u}_{1t} w_t^b}{\hat{\sigma}_1^b \sqrt{\sum_{t=1}^{T} x_t^2}},$$

where

$$\hat{\sigma}_1^{b2} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_{1t}^2 w_t^{b2} - \frac{1}{T} \left(\frac{1}{T} \sum_{t=1}^{T} x_t \hat{u}_{1t} w_t^b \right)^2 \left(\frac{1}{T^2} \sum_{t=1}^{T} x_t^2 \right)^{-1}.$$

Let $N_T^b = T^{-1} \sum_{t=1}^{T} x_t \hat{u}_{1t} w_t^b$. Conditionally on $\{y_t, x_t\}_{t=1}^T$, N_T^b is a Gaussian random variable with mean 0 and variance given by (12). Using Lemma 1 and (13), this implies that as $T \to \infty$,

$$N_T^b \xrightarrow{\mathcal{L}}_P \sigma_1 \sigma_2 \left(\int_0^1 X(r)^2 Q(r)^2 dr \right)^{1/2} Z,$$

where $Z \sim N(0,1)$, independent of $X(\cdot)$. This in turn implies that

$$\hat{\sigma}_1^{b2} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{1t}^2 w_t^{b2} + o_P(1) = \hat{\sigma}_1^2 + \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{1t}^2 (w_t^{b2} - 1) + o_P(1).$$

The limit of $\hat{\sigma}_1^2$ is given in (11); the second term has mean zero and variance $T^{-2}\sum_{t=1}^{T} 2E(\hat{u}_{1t}^4)$, which converges to zero as $T \to \infty$ because z_t has finite fourth moment. Therefore,

$$t_{ols}^b = \frac{N_T^b}{\hat{\sigma}_1^b \sqrt{T^{-2}\sum_{t=1}^{T} x_t^2}} \xrightarrow{\mathcal{L}}_P \left(\frac{\int_0^1 X(r)^2 Q(r)^2 dr}{\int_0^1 Q(r)^2 dr \int_0^1 X(r)^2 dr} \right)^{1/2} Z = \sqrt{A} Z.$$

□

We conclude this section with a small Monte Carlo experiment to investigate the effectiveness of both procedures in finite samples. Following KP08, we consider the model (1)–(3), with $T = 100$, $\sigma_1^2 = \sigma_2^2$, $z_t \sim$ i.i.d. $N(0, I_2)$, $a \in \{10, 0.01\}$, and $s \in \{0.1, 0.15, 0.2, 0.25, 0.3, 0.5, 0.7, 0.75, 0.8, 0.85, 0.9\}$. A bootstrap sample size of $B = 999$ has been used. Table 1 lists the Monte Carlo rejection frequencies (based on $100,000$ replications) at the nominal 5% level of t_{ols}, t_W and a wild bootstrap version of both tests (labelled t_{ols}^b and t_W^b, respectively). In agreement with KP08, we observe that the effect of a variance increase, in particular towards the end of the sample, is more pronounced than the effect of a variance decrease (which in turn is most pronounced at the beginning of the sample). The three methods to correct the size distortions are about equally successful in general; for the cases with the largest size distortions (a late positive shift), a combination of both approaches seems slightly more effective than the separate approaches.

Table 1 about here

4 Discussion

The results of this paper can be extended in various directions. First, the dimensions of both y_t and x_t could be larger than one, without qualitatively changing the results. Next, the restriction to a single and common (proportional) variance shift is inessential: all results will go through with minor notational change if $\Sigma_t = Q\left(\frac{t}{T}\right)^2 \Sigma$, with $Q(\cdot)$ a diagonal matrix function with càdlàg functions on the diagonal. An extension to allow for endogeneity and serial correlation in u_t is slightly less obvious, because time variation in (auto-) covariance matrices of u_t will typically also imply time variation of coefficient matrices in the fully modified OLS procedure, an (approximating) vector autoregression, or in a dynamic OLS regression. An extension of fully modified OLS to this situation, based on an estimate of the break point, has recently been proposed by Kourogenis *et al.* (2008). If one is prepared to assume a constant vector autoregression with only time variation in the innovation variance matrix, then the approach of Cavaliere *et al.* (2009) can be extended to inference on the cointegration parameters. We are currently investigating the effectiveness of heteroskedasticity-consistent variance matrix estimation in this framework, in comparison with the wild bootstrap.

Acknowledgments

Helpful comments from Giuseppe Cavaliere and Nikolaos Kourogenis are gratefully acknowledged.

References

Cavaliere G., Taylor, A.M.R., 2007. Testing for unit roots in time series models with nonstationary volatility. Journal of Econometrics 140, 919–947.

Cavaliere G., Taylor, A.M.R., 2008. Bootstrap unit root tests for time series with nonstationary volatility. Econometric Theory 24, 43–71.

Cavaliere, G., Rahbek, A., Taylor, A.M.R., 2009. Testing for cointegration in vector autoregressions with nonstationary volatility. Journal of Econometrics, forthcoming.

Giné, E., Zinn, J., 1990. Bootstrapping general empirical measures. Annals of Probability 18, 851–869.

Kourogenis, N., Pittis, N., 2008. Cointegration, variance shifts and the limiting distribution of the OLS estimator. Economics Letters 99, 103–106.

Kourogenis, N., Panopoulou, E., Pittis, N., 2008. Estimation and hypothesis testing of a cointegrating vector under a possible variance break. Working paper, Department of Banking and Financial Management, University of Piraeus.

Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Annals of Statistics 21, 255–285.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

Table 1: Rejection frequencies at nominal 5% level

s	0.1	0.15	0.2	0.25	0.3	0.5	0.7	0.75	0.8	0.85	0.9
						$a = 10$					
t_{ols}	0.065	0.073	0.078	0.086	0.092	0.127	0.162	0.167	0.166	0.161	0.152
t_W	0.059	0.062	0.060	0.061	0.061	0.066	0.068	0.069	0.068	0.068	0.068
t^b_{ols}	0.056	0.058	0.056	0.057	0.057	0.061	0.065	0.066	0.066	0.068	0.069
t^b_W	0.054	0.056	0.054	0.055	0.054	0.058	0.059	0.060	0.058	0.060	0.060
						$a = 0.01$					
t_{ols}	0.083	0.083	0.079	0.071	0.067	0.054	0.046	0.045	0.045	0.045	0.045
t_W	0.053	0.056	0.058	0.056	0.058	0.057	0.056	0.056	0.057	0.057	0.056
t^b_{ols}	0.053	0.053	0.056	0.054	0.055	0.054	0.054	0.053	0.054	0.054	0.053
t^b_W	0.058	0.057	0.060	0.057	0.057	0.055	0.054	0.053	0.053	0.054	0.053