COINTEGRATING RANK SELECTION IN MODELS WITH TIME-VARYING VARIANCE

By

Xu Cheng and Peter C.B. Phillips

COWLES FOUNDATION PAPER NO. 1362

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

2012

http://cowles.econ.yale.edu/

Journal of Econometrics 169 (2012) 155–165

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Cointegrating rank selection in models with time-varying variance[☆]

Xu Cheng[a,*], Peter C.B. Phillips[b,c,d]

[a] Department of Economics, University of Pennsylvania, United States
[b] Cowles Foundation, Yale University, United States
[c] University of Auckland, New Zealand
[d] Singapore Management University, Singapore

ARTICLE INFO

Article history:
Available online 25 January 2012

JEL classification:
C22
C32

Keywords:
Cointegrating rank
Heterogeneity
Information criteria
Model selection
Time varying variances

ABSTRACT

Reduced rank regression (RRR) models with time varying heterogeneity are considered. Standard information criteria for selecting cointegrating rank are shown to be weakly consistent in semiparametric RRR models in which the errors have general nonparametric short memory components and shifting volatility provided the penalty coefficient $C_n \to \infty$ and $C_n/n \to 0$ as $n \to \infty$. The AIC criterion is inconsistent and its limit distribution is given. The results extend those in Cheng and Phillips (2009a) and are useful in empirical work where structural breaks or time evolution in the error variances is present. An empirical application to exchange rate data is provided.

1. Introduction

Much attention has been given to econometric estimation and inferential procedures for time series with time-varying variances or nonstationary volatility. Among others, Pagan and Schwert (1990), Loretan and Phillips (1994), and Watson (1999) documented empirical evidence for temporal heterogeneity in the variation of many macroeconomic and financial time series. Particular concern has recently been given to the effect of the presence of heterogeneous unconditional variation and variance breaks on the validity of unit root tests. Several authors (Hamori and Tokihisa, 1997; Kim et al., 2002; Cavaliere, 2004; Cavaliere and Taylor, 2007) have shown that conventional unit root tests may suffer size distortions and reduced power when there is persistent heterogeneity in variation. Depending on the specific pattern of the volatility changes, the size distortions can be large enough to justify the use of more robust inferential techniques or adaptive estimation methods to secure gains in efficiency, such as those developed for autoregressive models (Phillips and Xu, 2006; Xu and Phillips, 2008). The effect of variance shifts on KPSS tests has also been studied (Busetti and Taylor, 2003; Cavaliere, 2004; Cavaliere and Taylor, 2005).

Modified unit root tests have been proposed to deal with various forms of departure from homoskedasticity for nonstationary time series. Kim and Phillips (2001) dealt with the case of a single abrupt change in variance by using a two-stage procedure where the breakpoint together with the pre- and post-break variances are estimated in the first step. Cavaliere and Taylor (2007) developed tests that are robust to multiple abrupt or smooth volatility changes using simulation based methods. And Beare (2007) used kernel methods to remove the heteroskedasticity before applying standard semiparametric procedures such as the Phillips–Perron test. Boswijk (2005) evaluated the power loss of various unit root tests, derived the asymptotic power envelope against a sequence of local alternatives to a unit root under nonstationary volatility and gave an adaptive test procedure based on volatility filtering.

In contrast to these univariate studies in the presence of persistent shifts in volatility, the present paper deals with multivariate systems and uses information based methods rather than Neyman Pearson tests. The focus of attention is the rank of the cointegrating space in a model with some unit roots. Analogous to scalar unit root tests, residual based cointegration tests suffer from size distortion under nonstationary volatility. Alternative methods based on vector autoregressions, such as the Johansen (1988, 1995) trace test, are also invalidated by time varying variances. Some of these methods impose strong parametric assumptions on the form of the model. The information theoretic approach taken

☆ We thank the Co-Editors and two anonymous referees for constructive comments. Cheng acknowledges the support of an Anderson Fellowship from the Cowles Foundation for Research in Economics. Phillips acknowledges partial support from the NSF under Grant No. SES 06-47086. The paper was completed while Cheng was a graduate student at Yale University.

* Corresponding author.
E-mail address: xucheng@sas.upenn.edu (X. Cheng).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.01.022

here uses a semiparametric framework and is shown to be robust to variance changes of a very general form. It may be used to consistently estimate cointegrating rank in a multivariate time series environment or as a scalar unit root test. In both cases, the procedure is robust to persistent shifts in volatility and is easy to implement in practical work.

The paper is closely related to past work on econometric model selection using information criteria. The most common applications of these methods involve choice of lag length in (vector) autoregression and variable choice in regression in parametric settings, but the methods have also been suggested for cointegrating rank selection (Phillips, 1996; Chao and Phillips, 1999). In particular, Cheng and Phillips (2009a, hereafter CP) show that cointegrating rank selection by suitable information criteria is consistent in a general semiparametric framework using reduced rank regression (RRR) within a simple VAR(1) model. In particular, RRR may be implemented without explicitly taking into account weak dependence in the errors. The present paper strengthens the results in CP by showing that these methods remain consistent when the errors are weakly dependent and there are persistent shifts in volatility. More specifically, information criteria are weakly consistent for selecting cointegrating rank provided that the penalty term goes to infinity at a rate slower than the sample size.

The approach is quite straightforward for practical implementation. Simulations indicate that under many forms of heteroskedasticity, the usual BIC criterion for cointegrating rank selection performs satisfactorily. The main practical import of the paper, therefore, is that the same cointegrating rank selection method may be used in empirical work for a wide range of semiparametric models of cointegration with shifting variances.

Another contribution of the paper is to provide a limit theory for regression in multivariable systems with some unit roots, weakly dependent errors and nonstationary volatility. This limit theory is useful in studying cases where reduced rank regressions are misspecified, possibly through the choice of inappropriate lag lengths in the vector autoregression or ignorance of the persistent shifts in variance.

The organization of the paper is as follows. Section 2 introduces the semiparametric heteroskedastic error correction model (ECM) and gives assumptions and estimation details. Asymptotic results are given in Section 3. Section 4 briefly reports some simulation results. An empirical application to exchange rate data is reported in Section 5. Section 6 concludes. Proofs and technical material are in the Appendix.

2. The semiparametric heteroskedastic ECM

As in CP and using the same notation, we consider the semiparametric ECM model

$$\Delta X_t = \alpha\beta' X_{t-1} + u_t, \quad t \in \{1, \ldots, n\}, \tag{1}$$

where X_t is an m-vector time series, and α and β are $m \times r_0$ full rank matrices. The integer r_0 is the unknown cointegrating rank parameter. The error term $\{u_t\}$ is weakly dependent and heterogeneously distributed according to

$$u_t = D(L)\varepsilon_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j},$$

$$\varepsilon_t = V(t/n)e_t, \quad e_t \sim iid\,(0, \Sigma_e), \tag{2}$$

where $V(\cdot) = \mathrm{diag}\{V_1(\cdot), \ldots, V_m(\cdot)\}$ and $V_k(\cdot)$, for $k = 1, \ldots, m$, is an unknown positive scale function. Under this specification, the innovation term ε_t has mean zero and time-varying variance $V(t/n)\Sigma_e V(t/n)$. The series X_t is initialized at $t = 0$ by some (possibly random) quantity $X_0 = O_p(1)$, although other initialization assumptions may be considered, as in Phillips (2008). Following conventions in the literature, we neglect the triangular array notation for $\{X_t\}$, $\{u_t\}$, and $\{\varepsilon_t\}$. The *iid* assumption in (2) is convenient but may be relaxed to allow for martingale difference errors, and hence conditional volatility in the errors u_t, provided the required functional limit theory given below in (6) still holds.

CP show that when the error term u_t in the semiparametric ECM (1) is stationary, i.e. $V(\cdot)$ is constant, the cointegrating rank r_0 can be consistently estimated by information criteria without explicitly taking into account the weak dependence structure of u_t. Specifically, the cointegrating rank is selected as

$$\widehat{r} = \arg\min_{0\le r\le m} IC(r), \quad \text{where}$$

$$IC(r) = \log|\widehat{\Sigma}(r)| + C_n n^{-1}(2mr - r^2), \tag{3}$$

$\hat{\Sigma}(r)$ is the residual covariance matrix obtained from RRR as if r were the true cointegrating rank, and C_n is a general penalty coefficient that may depend on the sample size. The coefficient $C_n = \log n, 2\log\log n$, or 2 corresponds to the BIC (Akaike, 1977, Rissanen, 1978, and Schwarz, 1978), Hannan and Quinn (1979) and Akaike (1973) penalties, respectively. Sample information-based versions of the coefficient C_n may also be employed, such as those in Wei's (1992) FIC criterion and Phillips and Ploberger's (1996) PIC criterion. The BIC version of (3) was given in Phillips and McFarland (1997). When u_t is stationary, weak consistency of this procedure only requires that C_n goes to infinity with sample size n but at a slower rate. This semi-parametric approach avoids possible misspecification on the weak dependent structure on u_t and is easy to implement in practice.

The current paper employs the same approach as (3) but allows for heterogeneous innovations and derives conditions on C_n to ensure consistent cointegrating rank selection. The information criterion approach has the advantage of being robust to weak dependence and time-varying variances in the errors, whereas variance heterogeneity is known to invalidate many commonly used approaches, such as residual based cointegration tests and VAR based methods (Johansen, 1988, 1995).

The following assumptions provide conditions on the weak dependence and time-varying structure of the error term u_t. Assumption 1 below imposes conditions on the linear process u_t that facilitate the partial sum limit theory. The 1-summability condition is helpful in validating (absolute) summability of the components in the BN decomposition. It is also used in the development of Cavaliere and Taylor (2007) for similar reasons, although it may be relaxed to 1/2 summability when it is used only to validate the functional limit theory, as here. Assumption 2 gives conditions on the innovation variance that are analogous to those used in Phillips and Xu (2006). These allow for a wide, but deterministic, evolution in the error variances. The variance function $V_k(\cdot)$ is defined on $(-\infty, 1]$ to allow the initialization to start from the infinite past. Interestingly, as the analysis below shows, persistent shifts in the variance of u_t does not mimic persistence in the observed series X_t, at least asymptotically. The explanation is that temporal evolution in volatility may occur without materially affecting unit root behavior in the individual series and co-movement across series. On the other hand, as in the observation of persistent series with noise, when the error variance in the noise is large enough, it can dominate finite sample behavior. Assumption 3 gives conditions that are standard in the study of reduced rank regressions with some unit roots (Johansen, 1988, 1995; Phillips, 1995).

Assumption 1. The lag polynomial $D(L) = \sum_{j=0}^{\infty} D_j L^j$ satisfies that $D_0 = I$, $D(1)$is full rank, and $\sum_{j=0}^{\infty} j||D_j|| < \infty$, where $||\cdot||$ is some matrix norm. The covariance matrix Σ_e is positive with unity diagonal elements and $E||e_t||^4 < \infty$.

Assumption 2. $V_k(\cdot)$, for $k = 1, \ldots, m$, is non-stochastic, measurable and uniformly bounded on the interval $(-\infty, 1]$, with a finite number of points of discontinuity, $V_k(\cdot) > 0$ and satisfies a Lipschitz condition except at points of discontinuity.

Assumption 3. (a) The determinantal equation $|I_m - (I_m + \alpha\beta')L| = 0$ has roots on or outside the unit circle, i.e. $|L| \geq 1$.
(b) Set $\Pi = I_m + \alpha\beta'$ where α and β are $m \times r_0$ matrices of full column rank r_0, $0 \leq r_0 \leq m$. (If $r_0 = 0$ then $\Pi = I_m$; if $r_0 = m$ then β has full rank m).
(c) The matrix $R = I_r + \beta'\alpha$ has eigenvalues within the unit circle.

For any given r, the information criterion $IC(r)$ in (3) involves a trade-off between fit and penalty, represented by $|\widehat{\Sigma}(r)|$ and C_n, respectively. Conditions on C_n that ensure consistency in the selection of cointegrating rank, therefore necessarily depend on the limit behavior of $|\widehat{\Sigma}(r)|$. Irrespective of the error structure, it is known from the algebra of RRR (Johansen, 1995) that

$$|\widehat{\Sigma}(r)| = |S_{00}|\Pi_{i=1}^{r}(1 - \widehat{\lambda}_i), \tag{4}$$

where $\widehat{\lambda}_i$, $1 \leq i \leq r$, are the r largest solutions to the usual canonical correlation equation involving the sample covariance matrix $S_{00} = n^{-1}\sum_{i=1}^{n}\Delta X_t \Delta X_t'$ and related matrices, as given by Johansen (1988, 1995) and CP.[1] The asymptotic properties of $|\widehat{\Sigma}(r)|$, $1 \leq r \leq m$, and therefore the conditions imposed on C_n, are determined by the limit behavior of $\widehat{\lambda}_i$ for $1 \leq i \leq m$. The limits of $\hat{\lambda}_i$, $1 \leq i \leq m$, are derived under *iid* and stationary errors by Johansen (1988, 1995) and CP, respectively. The next section derives the limit behavior of the roots $\widehat{\lambda}_i$ under the error structure (2), allowing for both weak dependence and unconditional heterogeneity, leading to the required conditions on the penalty coefficient C_n for consistency of the selection criteria.

3. Asymptotic results

To derive asymptotic results associated with the unit roots of the cointegrating system, we characterize the volatility of the k'th element of the innovation ε_t by its variance profile

$$\eta_k(s) := \left(\int_0^1 V_k(p)dp\right)^{-1}\int_0^s V_k(p)dp, \quad \text{and}$$
$$\sigma_k := \int_0^1 V_k(p)dp, \tag{5}$$

for $k = 1, \ldots, m$. The variance profile $\eta_k(s)$ is equal to s only when the innovation is homogeneous. The variance profile $\eta_k(s)$ is normalized by the average innovation variance σ_k so that $\eta_k(s)$ is an increasing homeomorphism on $[0, 1]$ with $\eta_k(0) = 0$ and $\eta_k(1) = 1$. The following lemma provides the building block for results associated with unit roots in the cointegrating system.

Lemma 1. *Under Assumptions 1–3,*
(a)

$$n^{-1/2}\sum_{s=1}^{[n\cdot]} u_s \Rightarrow B_V(\cdot), \quad \text{where } B_V(\cdot) = D(1)\int_0^{\cdot} V(s)dB_e(s) \tag{6}$$

and $B_e(\cdot)$ is a Brownian motion with variance Σ_e.
(b)

$$B_V(\cdot) = D(1)\Omega B_\eta(\cdot)\Sigma_e^{1/2},$$

where $\Omega = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$, $B_\eta(\cdot) = (B_1(\eta_1(\cdot)), \ldots, B_m(\eta_m(\cdot)))'$, and $B_1, \ldots, B_m$ are independent standard Brownian motions.

[1] Note that $|S_{00}|$ in $|\hat{\Sigma}(r)|$ is common across r and, hence, is irrelevant when minimizing $IC(r)$ across r.

These limit laws involve the variance transformed Brownian motion B_V, which is the Brownian motion under time deformation. In particular, at time $s \in [0, 1]$, $B_k(\eta_k(s))$ has the same distribution as standard Brownian motion B_k at time $\eta_k(s) \in [0, 1]$.

Corresponding to the stationary roots of the cointegrating system, the time-varying variance innovation produces a weak trend effect in sample mean convergence, as in Phillips and Xu (2006). In particular, a building block in the short memory asymptotics is the mean of the time-varying variance defined by

$$\overline{V} = \int_0^1 V(r)\Sigma_e V(r)'dr. \tag{7}$$

The next lemma gives the limit behavior of the roots $\{\hat{\lambda}_i, 1 \leq i \leq m\}$, employed in (4). We first define some of the notations used below. Employing the Wold representation as in CP and using the error structure (2), we write

$$\beta' X_t = G(L)\varepsilon_t = \sum_{j=0}^{\infty} G_j\varepsilon_{t-j} \quad \text{and}$$
$$\Delta X_t = W(L)\varepsilon_t = \sum_{j=0}^{\infty} W_j\varepsilon_{t-j},$$

where $G(L)$ and $W(L)$ are both lag polynomials whose coefficients can be determined via the partial sum (or generalized Granger) representation of X_t. Let $\alpha_\perp$ and $\beta_\perp$ be orthogonal complements to α and β, so that $[\alpha, \alpha_\perp]$ and $[\beta, \beta_\perp]$ are nonsingular and $\beta_\perp'\beta_\perp = I_{m-r}$.

Lemma 2. *Under Assumptions 1–3, when the true cointegration rank is r_0,*
(a) *$\widehat{\lambda}_i$ with $1 \leq i \leq r_0$, converge to the roots of*

$$|\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}| = 0, \tag{8}$$

where

$$\Sigma_{\beta\beta} = \sum_{j=0}^{\infty} G_j\overline{V}G_{j+h}', \qquad \Sigma_{00} = \sum_{j=0}^{\infty} W_j\overline{V}W_{j+h}', \quad \text{and}$$
$$\Sigma_{\beta 0} = \Sigma_{\beta 0}' = \sum_{j=0}^{\infty} G_j\overline{V}W_{j+h}'.$$

(b) *$\widehat{\lambda}_i$ with $r_0 + 1 \leq i \leq m$, decrease to zero at the rate n^{-1} and $\{n\widehat{\lambda}_i : i = r_0 + 1, \ldots, m\}$ converge weakly to the roots of*

$$\left|\rho\int_0^1 G_uG_u' - \left(\int_0^1 G_u dG_u'\beta_\perp' + \Psi\right)\widetilde{\alpha}_\perp(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp)^{-1}\widetilde{\alpha}_\perp' \right.$$
$$\left.\times\left(\beta_\perp\int_0^1 dG_uG_u' + \Psi'\right)\right| = 0, \tag{9}$$

where $G_u(r) = (\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' B_V(r)$ is $m - r_0$ dimensional variance transformed Brownian motion, $\widetilde{\alpha}_\perp$ is an $m \times (m - r)$ orthogonal complement to $\widetilde{\alpha} = \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}$ and $[\widetilde{\alpha}, \widetilde{\alpha}_\perp]$ is nonsingular, and

$$\Psi = \Psi_{wu}^1 + \Psi_{wv}\alpha', \quad \text{where}$$
$$\Psi_{wu}^1 = \sum_{h=1}^{\infty}\sum_{j=0}^{\infty}\beta_\perp' W_j\overline{V}D_{j+h}, \quad \text{and} \quad \Psi_{wv} = \sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\beta_\perp' W_j\overline{V}G_{j+h}.$$

Remarks. 1. Comparing Lemma 2 with corresponding results under homogeneous and martingale difference errors, we see that in all cases $\widehat{\lambda}_i$ with $1 \leq i \leq r_0$ are all positive in the limit and $\widehat{\lambda}_i$ with $r_0 + 1 \leq i \leq m$ converge to 0 at the rate n^{-1}. The broad limit behavior is therefore analogous in the general case.

2. However, under homogeneous innovations, $B_V(\cdot)$ becomes an m-vector Brownian motion with variance $\sigma^2 D(1)\Sigma_e D(1)'$ and $\overline{V}$ reduces to $\sigma^2\Sigma_e$, where $\sigma = \sigma_1 = \cdots = \sigma_m$. As such, the sample variance and covariance terms Σ_{00}, $\Sigma_{\beta\beta}$, $\Sigma_{\beta 0}$ and the one sided long run variance Ψ^1_{wu} and Ψ_{wv} all simplify to moments of ΔX_t and $\beta' X_t$, both of which are now stationary. Those results under homogeneous errors were given in CP.

Next we provide conditions on C_n so that the information criterion $IC(r)$ in (3) is weakly consistent for cointegrating rank selection. In order for the true cointegrating rank r_0 to be selected, C_n must be chosen such that $IC(r) - IC(r_0) > 0$ for any $r \neq r_0$ with probability approaching unity as $n \to \infty$. From (3) and (4),

$$IC\,(r) - IC\,(r_0) = \begin{cases} \sum_{i=r_0+1}^{r} \log\left(1-\widehat{\lambda}_i\right) + C_n n^{-1} K\,(r) & \text{for } r > r_0 \\ -\sum_{i=r+1}^{r_0} \log\left(1-\widehat{\lambda}_i\right) + C_n n^{-1} K\,(r) & \text{for } r < r_0, \end{cases}$$

where $K\,(r) = (r - r_0)\,(2m - r - r_0) = (r - r_0)\,(2m - 2r_0 - (r - r_0))$. Observe that $K\,(r)$ is positive over $r \in (r_0, 2m - r_0)$. Thus, for $r > r_0$, $\sum_{i=r_0+1}^{r} \log(1-\widehat{\lambda}_i) = O_p\left(n^{-1}\right)$ as $n \to \infty$ by Lemma 2(b), $K\,(r) > 0$ as just noted, and so provided $C_n \to \infty$ the second term dominates when $r > r_0$ and $IC\,(r) > IC\,(r_0)$ with probability approaching unity as $n \to \infty$. For $r < r_0$, $-\sum_{i=r+1}^{r_0} \log(1-\widehat{\lambda}_i)$ converges to a finite positive value by virtue of Lemma 2(a) and $C_n n^{-1} \to 0$ provided $C_n = o\,(n)$ so that again $IC\,(r) > IC\,(r_0)$. Hence, $IC\,(r) - IC\,(r_0) > 0$ as $n \to \infty$ for all $r \neq r_0$. In sum, $IC\,(r)$ is a consistent rank selector if $C_n \to \infty$ at a rate slower than n, even in the presence of weak dependence and time-varying variances in the errors.

A key observation from the above analysis is that the determining factor for the conditions on C_n for consistency is the order of the $\widehat{\lambda}_i$, $1 \le i \le m$, rather than their exact limits. For example, we only need that the $\widehat{\lambda}_i$, $r_0 + 1 \le i \le m$, decrease to zero at the rate n^{-1} to get the given restriction on C_n, whereas the precise limiting form of the scaled roots $n\widehat{\lambda}_i$ defined in (9) does not influence consistency. Of course, the finite-sample behavior of the selection procedure will be affected by the distribution of these roots.

Since the asymptotic orders of $\widehat{\lambda}_i$, $1 \le i \le m$, are unaffected by heterogeneity in the errors, the information criterion $IC(r)$ proposed in CP for stationary errors remains valid under heterogeneity, although the finite-sample properties of the method are likely to be influenced by the form of the variance profile of the innovations. The asymptotic results are presented in Theorem 1 below and finite-sample behavior is investigated by simulations reported in Section 4.

Theorem 1. *Under Assumptions 1–3,*

(a) *the criterion IC(r) is weakly consistent for selecting the rank of cointegration provided $C_n \to \infty$ at a slower rate than n;*
(b) *the asymptotic distribution of the AIC criterion (IC(r) with coefficient $C_n = 2$) is given by*

$$\lim_{n\to\infty} P(\hat{r}_{AIC} = r_0) = P\left[\bigcap_{r=r_0+1}^{m}\left\{\sum_{i=r_0+1}^{r} \xi_i < 2(r-r_0)(2m-r-r_0)\right\}\right],$$

$$\lim_{n\to\infty} P(\hat{r}_{AIC} = r | r < r_0) = 0, \quad \textit{and}$$

$$\lim_{n\to\infty} P(\hat{r}_{AIC} = r | r > r_0) = P\left\{\left(\bigcap_{r'=r+1}^{m}\left\{\sum_{i=r+1}^{r'} \xi_i < 2(r'-r)(2m-r'-r)\right\}\right) \cap \left(\bigcap_{r'=r_0}^{r-1}\left\{\sum_{i=r'+1}^{r} \xi_i > 2(r-r')(2m-r-r')\right\}\right)\right\},$$

where $\xi_{r_0+1}, \ldots, \xi_m$ are the ordered roots of the limiting determinantal equation (9).

This result provides a convenient basis for consistent cointegration rank selection in most empirical contexts under very general assumptions on the errors. As in the homogeneous variance case, BIC, HQ and other information criteria with $C_n \to \infty$ and $C_n/n \to 0$ are all consistent in the presence of weakly dependent errors with time-varying variance. The information criterion consistently selects cointegrating rank under general assumptions on the errors without having to specify any parametric model of short memory or heterogeneity. When $m = 1$, the unit root model corresponds to $r_0 = 0$ and $r_0 = 1$ to the stationary model. Unlike some standard unit root tests, model choice by information criteria is then robust to the presence of permanent shifts in variance. The theorem also applies in the case of models with intercepts and drifts. Generalizations are also possible to cases where some of the time series are near integrated.

AIC is inconsistent, asymptotically never underestimates cointegrating rank, and favors more liberally parameterized systems. This outcome is analogous to the well-known overestimation tendency of AIC in lag length selection in autoregression and is consistent with earlier results on cointegration rank selection under homogeneous errors.

4. Simulations

This section reports some brief simulations for different forms of the variance function $V(\cdot)$, different settings for the true cointegrating rank, and various choices of the penalty coefficient C_n. The data generating process follows (1) and the design of the reduced rank coefficient follows CP. Thus, when $r_0 = 0$ we have $\alpha'\beta = 0$, and when $r_0 = 1$ the reduced rank coefficient matrix is set to

$$\alpha'\beta = (1, 0.5)\begin{pmatrix}-1\\1\end{pmatrix}.$$

When $r_0 = 2$, two different simulations (design A and design B) were performed, one with smaller and one with larger stationary roots as follows:

$$A: \alpha'\beta = \begin{pmatrix}-0.5 & 0.1\\ 0.2 & -0.4\end{pmatrix},$$

with stationary roots $\lambda_i[I + \beta'\alpha] = \{0.7, 0.4\}$;

$$B: \alpha'\beta = \begin{pmatrix}-0.5 & 0.1\\ 0.2 & -0.15\end{pmatrix},$$

with stationary roots $\lambda_i[I + \beta'\alpha] = \{0.9, 0.45\}$.

Following Cavaliere and Taylor (2007), we assessed the performance of the information criteria uncontaminated by serial dependence by setting $u_t = \varepsilon_t$. To evaluate the method under weak dependence, simulations were also conducted under the following AR(1), MA(1), ARMA(1,1) formulations

$$u_t = Au_{t-1} + \varepsilon_t, \qquad u_t = \varepsilon_t + B\varepsilon_{t-1}, \qquad u_t = Au_{t-1} + \varepsilon_t + B\varepsilon_{t-1}, \tag{10}$$

with coefficient matrices $A = \psi I_m$, $B = \phi I_m$, where $|\psi| < 1$, $|\phi| < 1$. The innovations with time-varying variance are

$$\varepsilon_t = V(t/n)e_t \quad \text{and} \quad e_t = iid\,N(0, \Sigma_\varepsilon), \tag{11}$$

where

$$\Sigma_\varepsilon = \begin{pmatrix} 1+\theta & 0 \\ 0 & 1-\theta \end{pmatrix} > 0.$$

The parameters for these models were set to $\psi = \phi = 0.4$ and $\theta = 0.25$.

The design of the variance matrix $V(\cdot)$ follows that in Cavaliere (2004), Cavaliere and Taylor (2007), and Phillips and Xu (2006). We assume that for any $s \in (-\infty, 1]$, the $m \times m$ diagonal variance matrix $V(s) = g(s)I_m$, where $g(\cdot)$ is a real positive function. Under this setup, all variables share the same variance profile, characterized by the variance function $g(\cdot)$. Three models for the variance function $g(\cdot)$ were used:

$$\begin{aligned} &1.\ g(s) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)1_{\{s \ge \tau\}}, \quad s \in [0,1], \\ &2.\ g(s) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)1_{\{\tau \le s < 1-\tau\}}, \\ &\quad s \in [0,1],\ \tau \in [0, 1/2], \\ &3.\ g(s) = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)s^m, \quad s \in [0,1]. \end{aligned} \tag{12}$$

There is a single volatility shift from σ_0^2 to σ_1^2 at time $[\tau n]$ in model 1 and there are two volatility shifts in model 2, which happens at time $[\tau n]$ and $1 - [\tau n]$, respectively. In contrast to the abrupt volatility jumps in these two models, model 3 models the situation where volatility changes smoothly from σ_0^2 to σ_1^2. The parameters in the simulation are set up as follows. In model 1, the break date τ takes values within the set $\{0.1, 0.5, 0.9\}$, so that early, middle and late breaks are all investigated. In model 2, τ takes value from $\{0.1, 0.4\}$, where a small τ corresponds to the case where the first jump happens early in the sample and the second jump happens late in the sample. In model 3, we allow for both linear trend and quadratic trend by setting $m \in \{1, 2\}$. Without loss of generality, we set $\sigma_0^2 = 1$ in all cases. The steepness of the break is measured by the ratio of the post-break and pre-break standard deviation: $\delta = \sigma_1/\sigma_0$, which takes values within the set $\{0.2, 5\}$ for all three models to allow for both positive ($\delta > 1$) and negative ($\delta < 1$) shifts. The performance of AIC and BIC[2] was investigated for sample sizes $n = 100, 400$ in all cases including 50 additional observations to eliminate start-up effects from the initializations $X_0 = 0$ and $\varepsilon_0 = 0$. The results are based on 20, 000 replications.

Tables 1–3 give simulation results for design A where the error u_t follows an AR(1) process. Similar results were obtained for the other error generating schemes in (10). As is evident in the tables, BIC generally performs well under different forms of volatility changes when the true rank r_0 is 1 or 2, although when $r_0 = 0$, it may overestimate in some cases under abrupt volatility shifts, depending on the pattern of the changes. Specifically, in model 1, the overestimation tends to happen when there is an early negative shift ($\tau = 0.1, \delta = 0.2$) or a late positive ($\tau = 0.9, \delta = 5$) shift, but not under early positive shifts or late negative shifts; in model 2, the overestimation happens when a very early shift is positive and a very late shift is negative. In the worst case, BIC selects the true cointegration rank $r_0 = 0$ with a probability around 65% when the sample size is 400. We also observe that the conventional tendency of BIC to underestimate order (here cointegrating rank) is mild when $n = 100$ and disappears completely when $n = 400$. These results are analogous to those in Phillips and Xu (2006), who show that in a stable autoregressive model various t statistics tend to over-reject under early negative shifts or late positive shifts and that this tendency is attenuated when the error variance dynamics follow a polynomial shape as in model 3. In all cases, BIC performs much more satisfactorily than AIC, which has a strong tendency to overestimate order, just as it does in lag length selection in autoregressive models.

Table 1
Cointegration rank selection in design A when u_t follows an AR(1) process under model 1.

τ	δ	$\hat{r}$	$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
0.1	0.2	0	0.24	0.65	0.00	0.00	0.00	0.00	0.21	0.50	0.00	0.00	0.00	0.00
		1	0.65	0.34	0.75	0.91	0.00	0.00	0.65	0.47	0.76	0.88	0.00	0.06
		2	0.12	0.01	0.25	0.09	1.00	1.00	0.14	0.03	0.24	0.12	1.00	0.94
	1	0	0.52	0.95	0.00	0.00	0.00	0.00	0.48	0.88	0.00	0.00	0.00	0.01
		1	0.37	0.04	0.77	0.97	0.00	0.00	0.40	0.11	0.78	0.94	0.00	0.05
		2	0.11	0.00	0.23	0.03	1.00	1.00	0.12	0.02	0.22	0.06	1.00	0.94
	5	0	0.51	0.93	0.00	0.00	0.00	0.00	0.49	0.87	0.00	0.00	0.00	0.01
		1	0.38	0.06	0.73	0.95	0.00	0.00	0.39	0.12	0.73	0.92	0.00	0.08
		2	0.11	0.01	0.27	0.05	1.00	1.00	0.12	0.02	0.27	0.08	1.00	0.92
0.5	0.2	0	0.38	0.87	0.00	0.00	0.00	0.00	0.33	0.73	0.00	0.00	0.00	0.01
		1	0.53	0.13	0.76	0.94	0.00	0.00	0.56	0.26	0.77	0.91	0.00	0.05
		2	0.09	0.00	0.24	0.06	1.00	1.00	0.11	0.02	0.23	0.09	1.00	0.95
	5	0	0.34	0.81	0.00	0.00	0.00	0.00	0.36	0.73	0.00	0.00	0.00	0.02
		1	0.45	0.17	0.62	0.90	0.00	0.00	0.45	0.23	0.64	0.87	0.02	0.13
		2	0.21	0.03	0.38	0.10	1.00	1.00	0.19	0.04	0.36	0.13	0.98	0.85
0.9	0.2	0	0.53	0.95	0.00	0.00	0.00	0.00	0.48	0.88	0.00	0.00	0.00	0.00
		1	0.37	0.04	0.77	0.96	0.00	0.00	0.41	0.11	0.79	0.94	0.00	0.03
		2	0.09	0.00	0.23	0.04	1.00	1.00	0.11	0.01	0.21	0.06	1.00	0.97
	5	0	0.27	0.64	0.00	0.00	0.00	0.00	0.30	0.61	0.00	0.00	0.00	0.03
		1	0.52	0.32	0.60	0.87	0.00	0.01	0.51	0.34	0.67	0.87	0.16	0.30
		2	0.22	0.04	0.40	0.13	1.00	0.99	0.19	0.04	0.33	0.13	0.84	0.67

Table 2
Cointegration rank selection in design A when u_t follows an AR(1) process under model 2.

τ	δ	$\hat{r}$	$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
0.1	0.2	0	0.50	0.93	0.00	0.00	0.00	0.00	0.50	0.87	0.00	0.00	0.00	0.01
		1	0.38	0.06	0.74	0.96	0.00	0.00	0.38	0.12	0.74	0.92	0.00	0.08
		2	0.11	0.01	0.26	0.04	1.00	1.00	0.12	0.02	0.26	0.08	1.00	0.92
	5	0	0.24	0.65	0.00	0.00	0.00	0.00	0.21	0.50	0.00	0.00	0.00	0.00
		1	0.64	0.34	0.75	0.90	0.00	0.00	0.65	0.48	0.76	0.89	0.00	0.05
		2	0.12	0.01	0.25	0.10	1.00	1.00	0.14	0.02	0.24	0.11	1.00	0.94
0.4	0.2	0	0.38	0.84	0.00	0.00	0.00	0.00	0.39	0.77	0.00	0.00	0.00	0.02
		1	0.44	0.14	0.65	0.92	0.00	0.00	0.44	0.20	0.67	0.89	0.01	0.12
		2	0.18	0.02	0.35	0.08	1.00	1.00	0.17	0.04	0.33	0.11	0.99	0.87
	5	0	0.33	0.82	0.00	0.00	0.00	0.00	0.27	0.65	0.00	0.00	0.00	0.01
		1	0.57	0.17	0.75	0.93	0.00	0.00	0.62	0.33	0.76	0.90	0.00	0.05
		2	0.09	0.01	0.25	0.07	1.00	1.00	0.11	0.02	0.24	0.10	1.00	0.94

Table 3
Cointegration rank selection in design A when u_t follows an AR(1) process under model 3.

m	δ	$\hat{r}$	$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
1.	0.2	0	0.44	0.90	0.00	0.00	0.00	0.00	0.37	0.76	0.00	0.00	0.00	0.00
		1	0.46	0.09	0.78	0.95	0.00	0.00	0.52	0.23	0.78	0.92	0.00	0.03
		2	0.10	0.00	0.22	0.05	1.00	1.00	0.12	0.02	0.22	0.08	1.00	0.96
	5	0	0.41	0.86	0.00	0.00	0.00	0.00	0.40	0.77	0.00	0.00	0.00	0.02
		1	0.43	0.13	0.66	0.93	0.00	0.00	0.43	0.20	0.66	0.88	0.02	0.14
		2	0.16	0.02	0.34	0.07	1.00	1.00	0.17	0.03	0.34	0.12	0.98	0.84
2.	0.2	0	0.49	0.93	0.00	0.00	0.00	0.00	0.43	0.82	0.00	0.00	0.00	0.00
		1	0.42	0.07	0.78	0.96	0.00	0.00	0.46	0.16	0.78	0.93	0.00	0.03
		2	0.09	0.00	0.22	0.04	1.00	1.00	0.11	0.01	0.22	0.07	1.00	0.97
	5	0	0.37	0.79	0.00	0.00	0.00	0.00	0.40	0.74	0.00	0.00	0.00	0.03
		1	0.45	0.18	0.63	0.92	0.00	0.00	0.43	0.23	0.68	0.90	0.05	0.19
		2	0.18	0.03	0.37	0.08	1.00	1.00	0.17	0.03	0.32	0.10	0.95	0.78

To show the effect of variance shifts when they are uncontaminated by temporal dependence, we performed simulations for design A under independent errors with the variance structure specified in (12). To save space, we only report the results under model 1, as shown in Table 4. Comparing Tables 1 and 4, we find

[2] It is shown in CP that AIC and BIC generally have better performance than other criteria such as Hannan–Quinn (HQ) or criteria with even weaker penalties than HQ such as $C_n = \log\log\log n$.

Table 4
Cointegration rank selection in design A when u_t is independent under model 1.

τ	δ	$\widehat{r}$	$n=400$						$n=100$					
			$r_0=0$		$r_0=1$		$r_0=2$		$r_0=0$		$r_0=1$		$r_0=2$	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
0.1	0.2	0	0.49	0.93	0.00	0.00	0.00	0.00	0.43	0.82	0.00	0.00	0.00	0.00
		1	0.42	0.07	0.78	0.96	0.00	0.00	0.46	0.16	0.78	0.93	0.00	0.03
		2	0.09	0.00	0.22	0.04	1.00	1.00	0.11	0.01	0.22	0.07	1.00	0.97
	1	0	0.52	0.95	0.00	0.00	0.00	0.00	0.64	0.98	0.00	0.00	0.00	0.00
		1	0.37	0.04	0.77	0.97	0.00	0.00	0.31	0.02	0.81	0.96	0.00	0.00
		2	0.11	0.00	0.23	0.03	1.00	1.00	0.05	0.00	0.19	0.04	1.00	1.00
	5	0	0.37	0.79	0.00	0.00	0.00	0.00	0.40	0.74	0.00	0.00	0.00	0.03
		1	0.45	0.18	0.63	0.92	0.00	0.00	0.43	0.23	0.68	0.90	0.05	0.19
		2	0.18	0.03	0.37	0.08	1.00	1.00	0.17	0.03	0.32	0.10	0.95	0.78
0.5	0.2	0	0.48	0.93	0.00	0.00	0.00	0.00	0.43	0.83	0.00	0.00	0.00	0.00
		1	0.43	0.07	0.77	0.96	0.00	0.00	0.46	0.16	0.78	0.93	0.00	0.03
		2	0.09	0.00	0.23	0.04	1.00	1.00	0.11	0.01	0.22	0.07	1.00	0.96
	5	0	0.37	0.80	0.00	0.00	0.00	0.00	0.40	0.75	0.00	0.00	0.00	0.03
		1	0.45	0.18	0.63	0.91	0.00	0.00	0.44	0.22	0.69	0.90	0.06	0.19
		2	0.19	0.03	0.37	0.09	1.00	1.00	0.16	0.03	0.31	0.10	0.94	0.78
0.9	0.2	0	0.49	0.93	0.00	0.00	0.00	0.00	0.42	0.83	0.00	0.00	0.00	0.00
		1	0.42	0.07	0.77	0.96	0.00	0.00	0.46	0.16	0.78	0.93	0.00	0.03
		2	0.09	0.00	0.23	0.04	1.00	1.00	0.11	0.01	0.22	0.07	1.00	0.96
	5	0	0.36	0.80	0.00	0.00	0.00	0.00	0.40	0.75	0.00	0.00	0.00	0.03
		1	0.45	0.18	0.63	0.91	0.00	0.00	0.44	0.22	0.68	0.90	0.05	0.20
		2	0.19	0.03	0.37	0.09	1.00	1.00	0.16	0.03	0.32	0.10	0.95	0.77

Table 5
Cointegration rank selection in design B when u_t follows an AR(1) process under model 1.

τ	δ	$\widehat{r}$	$n=400$						$n=100$					
			$r_0=0$		$r_0=1$		$r_0=2$		$r_0=0$		$r_0=1$		$r_0=2$	
			AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
0.1	0.2	0	0.24	0.65	0.00	0.00	0.00	0.00	0.21	0.50	0.00	0.00	0.00	0.02
		1	0.65	0.34	0.75	0.91	0.00	0.05	0.65	0.47	0.76	0.88	0.09	0.36
		2	0.12	0.01	0.25	0.09	1.00	0.95	0.14	0.03	0.24	0.12	0.91	0.62
	1	0	0.52	0.95	0.00	0.00	0.00	0.00	0.48	0.88	0.00	0.00	0.00	0.05
		1	0.37	0.04	0.77	0.97	0.00	0.02	0.40	0.11	0.78	0.94	0.26	0.75
		2	0.11	0.00	0.23	0.03	1.00	0.98	0.12	0.02	0.22	0.06	0.74	0.21
	5	0	0.51	0.93	0.00	0.00	0.00	0.00	0.49	0.87	0.00	0.00	0.00	0.06
		1	0.38	0.06	0.73	0.95	0.00	0.03	0.39	0.12	0.73	0.92	0.32	0.78
		2	0.11	0.01	0.27	0.05	1.00	0.97	0.12	0.02	0.27	0.08	0.68	0.16
0.5	0.2	0	0.38	0.87	0.00	0.00	0.00	0.00	0.33	0.73	0.00	0.00	0.00	0.04
		1	0.53	0.13	0.76	0.94	0.00	0.03	0.56	0.26	0.77	0.91	0.15	0.57
		2	0.09	0.00	0.24	0.06	1.00	0.97	0.11	0.02	0.23	0.09	0.85	0.39
	5	0	0.34	0.81	0.00	0.00	0.00	0.00	0.36	0.73	0.00	0.00	0.00	0.08
		1	0.45	0.17	0.62	0.90	0.00	0.09	0.45	0.23	0.64	0.87	0.36	0.69
		2	0.21	0.03	0.38	0.10	1.00	0.92	0.19	0.04	0.36	0.13	0.64	0.23
0.9	0.2	0	0.53	0.95	0.00	0.00	0.00	0.00	0.48	0.88	0.00	0.00	0.00	0.04
		1	0.37	0.04	0.77	0.96	0.00	0.01	0.41	0.11	0.79	0.94	0.16	0.70
		2	0.09	0.00	0.23	0.04	1.00	0.99	0.11	0.01	0.21	0.06	0.84	0.26
	5	0	0.27	0.64	0.00	0.00	0.00	0.00	0.30	0.61	0.00	0.00	0.00	0.08
		1	0.52	0.32	0.60	0.87	0.08	0.25	0.51	0.34	0.67	0.87	0.38	0.59
		2	0.22	0.04	0.40	0.13	0.92	0.75	0.19	0.04	0.33	0.13	0.62	0.33

that BIC is generally more reliable when the errors have low temporal dependence. Similar results were found for models 2 and 3.

The results for design B, where the stationary roots of the system are closer to unity, are shown in Table 5. Just as in CP, when the stationary root is large, BIC has a tendency to underestimate the rank when $n = 100$ and $r_0 = 2$, thereby choosing more parsimoniously parameterized system in this case. When $n = 400$, the underestimation is significantly attenuated.

To distinguish the effect of time-varying variance on the performance of AIC and BIC, we report results with $\delta = 1$, covering the constant variance profile case, as benchmarks in Tables 1, 4 and 5. Comparing outcomes under heterogeneous variances with these benchmarks, it is apparent that the presence of heterogeneity impacts the performance of both AIC and BIC. But BIC continues to outperforms AIC in general, as under homogeneous errors.

One situation where AIC does perform better than BIC is when the true cointegrating rank r_0 is equal to m (or is close to m when m is large). The reason is that AIC has a tendency to overestimate, as is clear from Theorem 1. However, in the rank selection context, there is an upper bound, m, of the choice set, so that it is impossible to overestimate when the true rank is r_0, which gives AIC a slight advantage in this case. The upper bound is therefore a major difference between cointegrating rank selection and lag order selection in an autoregression. In spite of this aspect of AIC, BIC is still the preferred criteria based on its overall superior performance.

In summary, the simulation results show that the BIC criterion for cointegration rank selection is robust to weak dependence and heterogeneity of the errors, generally confirming the asymptotic theory. The main weakness of BIC is that it tends to overestimate when early negative or late positive volatility shifts happen in a system without cointegration and to underestimate when the system is stationary but with a root near unity. The performance of BIC significantly improves as the sample size gets larger, the volatility shifts become smoother, or the temporal dependence of the errors is weaker. In all cases, BIC performs much better than alternative criteria such as AIC and seems sufficiently reliable to recommend for empirical practice.

5. Empirical application

This section reports the application of model selection techniques to cointegrating rank estimation in a dynamic exchange rate system. Using Johansen's trace test, Baillie and Bollerslev (1989) found evidence of one cointegration relation in vector autoregressions of seven daily spot and seven one-month forward rates. They concluded that these floating exchange rates follow one long-run equilibrium path. However, when adding an intercept to the model, Diebold et al. (1994) found no support for a cointegrating relation in these data. In addition to conventional cointegration tests, various fractional cointegration formulations have been considered in the same dynamic exchange rate setting, including Baillie and Bollerslev (1994), Kim and Phillips (2001), Nielsen (2004), Hassler et al. (2006), and Nielsen and Shimotsu (2007). These papers on fractional cointegration generally agree on the existence of fractional cointegration among the exchange rates of different currencies under the floating exchange rate regime.

The focus of the present application is the use of semiparametric rank selection methods to investigate possible cointegrating relations among exchange rates. It is now a well-established stylized fact that many macro-economic and financial variables, including exchange rates, are characterized by breaks in volatility. Our approach, with its robustness to shifting variances including both abrupt breaks and smooth transitions, seems well suited to this application. Moreover, there is no need to specify a particular parametric model for variance shifts or weak dependence in our approach, making it easy to implement and robust to a variety of different model specifications.

Our base data set concentrates on the same exchange rates as those studied in the literature cited above. This data comprised log exchange rates for seven currencies: the Canadian Dollar, French Franc, Deutsche Mark, Italian Lira, Japanese Yen, Swiss Franc and British Pound, all relative to the US Dollar. Baillie and Bollerslev (1989, 1994) and Diebold et al. (1994) used these seven nominal exchange rates observed daily from 1980 to 1985, Kim and Phillips (2001) used quarterly data from 1957 to 1997, and Nielsen and Shimotsu (2007) applied their estimation techniques to a data set of monthly averages of noon (EST) buying rates running from 1974:1 through 2001:12. Our base data (taken from the Board of Governors of the Federal Reserve System) also involves monthly averages of noon buying rates and runs from 1973:4 to 2008:12. These data also include exchange rate data for the Euro from 1999:1 to 2008:12. To complement this series, we constructed

Table 6
Mean and standard deviation (SD) of exchange rate returns.

Mean	Pound	0.001	Yen	0.0005	Canadian	−0.0023	Euro	−0.0002
SD		0.0248		0.0138		0.0276		0.0256

from the base data a pre 1999:1 time series for the Euro using a simple average of the Deutsche Mark, French Franc, and Italian Lira. The empirical results reported below are for the British Pound, Euro, Yen, and Canadian Dollar.[3] To illustrate the volatilities, the first order differences are plotted in Fig. 1 and their descriptive statistics are presented in Table 6.

The time-varying behavior of the exchange rate volatilities is well characterized by its variance profile $\eta_k(s)$, for $k = 1, \ldots, m$, which is increasing from 0 to 1 and only equal to s under homogeneous errors. We first estimate the variance profile of each exchange rate series using the method of Cavaliere and Taylor (2007). Let $X_{t,k}$ be the k'th element of X_t and $\{\widehat{u}_{t,k}\}$ denote the residuals from the linear regression of $\widehat{X}_{t,k}$ on $\widehat{X}_{(t-1),k}$, where $\widehat{X}_{t,k}$ is the residual of $X_{t,k}$ after detrending.[4] Detrending $X_{t,k}$ is necessary when we include an intercept in (1). The estimator of the variance profile, which is the sample analogue of (5) linearly interpolated between the observed sample data, can be written as

$$\widehat{\eta}_k(s) = \frac{\sum_{t=1}^{[ns]} \widehat{u}_{t,k}^2 + (ns - [ns])\widehat{u}_{([ns]+1),k}^2}{\sum_{t=1}^{n} \widehat{u}_{t,k}^2}. \tag{13}$$

Cavaliere and Taylor (2007) show that $\widehat{\eta}_k(\cdot)$ is a uniformly consistent estimator for the variance profile $\eta_k(\cdot)$.

The estimated variance profiles are presented in Fig. 2. In Fig. 2, the 45° line corresponds to the variance profile for homogeneous errors. During the relatively long time span after 1973, we see that none of the series experienced particularly sharp changes in volatility, although there is evidence of smooth transitions in some cases. Specifically, in this period, the Canadian Dollar has a smooth volatility profile, with a steady volatility during much of the period followed by a increase in volatility over recent years. The British Pound has a more abrupt positive shift in volatility in the 1980s followed by a subsequent negative shift. The variance profiles of the Yen and Euro exhibit only small shifts in volatility.

The cusum of squares test is conducted to test the null hypothesis that $u_{t,k}$, the error in the AR(1) process $X_{t,k}$, is covariance stationary for $k = 1, \ldots, m$.[5] To introduce the test statistic, now we use u_t to denote $u_{t,k}$ for simplicity. Following Pagan and Schwert (1990) and Loretan and Phillips (1994), the test statistic of the cusum of squares test is

$$\psi(s) = (n\hat{\nu})^{1/2} \sum_{j=1}^{[ns]} (u_j^2 - \tilde{\mu}_2), \quad \text{where } \tilde{\mu}_2 = n^{-1} \sum_{i=1}^{n} u_t^2,$$

and $\hat{\nu}$ is an consistent estimation of the long run variance of $\{u_t^2, 1 \le t \le n\}$ as in Phillips (1987). Under the null,

$$\psi(s) \to_d B(s),$$

where $B(s)$ is a standard Brownian Bridge on $C[0, 1]$. Fig. 3 plots the 95% confidence band based on $B(s)$ and the sample path of $\psi(s)$ for each currency. The cusum of squares tests show a strong rejection of variance constancy in the case of the British Pound and the Canadian Dollar and only marginal rejections of constancy in the case of the Yen and the Euro,[6] which confirm the descriptive characteristics of the variance profile in Fig. 2.

Table 7
Cointegration rank estimation.[8]

r	0	1	2	3	4	$\widehat{r}$
BIC	0.000	0.054	0.091	0.127	0.140	**0**
AIC	0.000	−0.012	−0.022	−0.014	−0.011	**2**

Information criteria are first used to reveal the dominant time series characteristic of the exchange rate data with $r = 0$ signifying $I(1)$ and $r = 1$ signifying $I(0)$. Both BIC and Phillips–Perron unit root tests show all series to be I(1) processes. The findings confirm earlier conclusions that nominal exchange rates are well characterized as $I(1)$ processes (c.f. Corbae and Ouliaris, 1988; Baillie and Bollerslev, 1989).[7]

Next, cointegrating rank among the four exchange rates is estimated by AIC and BIC under (1). The method allows for both weak dependence and variance heterogeneity as detected in Figs. 2 and 3. The estimation results are presented in Table 7. AIC finds 2 cointegrating relations and BIC finds no cointegration in the system. Considering the overestimation problem associated with AIC and the small underestimation probability of BIC given our large sample size, we conclude that there is no $I(1)/I(0)$ cointegration in the exchange rate dynamic system. Johansen's trace test also suggests no cointegration among these four series. Our result is consistent with the findings in Diebold et al. (1994), where Johansen trace test and seven currencies as in our base data were used. Compared with Johansen's method, our procedure does not require a first step estimation of the number of lags in the ECM, is more robust to model specification and is valid in the presence of time-varying variance.

6. Conclusion

This paper shows that cointegrating rank can be consistently selected by information criteria under weak conditions on the expansion rate of the penalty coefficient. In contrast to traditional reduced rank and other cointegration estimation methodologies, our method does not require a full parametric model and it is robust to both weak dependence and variance heterogeneity. As a cointegrating rank selector or as a simple unit root test it offers substantial convenience to the empirical researcher in the presence of these complications.

Some further extensions of this semiparametric cointegrating rank selection approach are possible and may be useful in empirical research. We mention a few ideas here. First, allowance

[3] Empirical results for the larger data set of individual currencies up to 1999:1 are reported in the original version of this paper (Cheng and Phillips, 2009b). These results are broadly similar to those reported here over the longer period that include the Euro.

[4] Estimation and tests based on $\hat{u}_t$ obtained under the unit root null hypothesis produced similar results.

[5] Tests are conducted under both the null hypothesis that $X_{t,k}$ are unit root processes for $k = 1, \ldots, m$ and without the hypothesis. The statistical results are similar for both cases.

[6] We can also use $\inf_s(\psi(s))$ and $\sup_s(\psi(s))$ as test statistics. Their asymptotic properties and critical values are available in Loretan and Phillips (1994). The British Pound rejects covariance constancy at 10% level with the sup test and the Canadian Dollar rejects covariance constancy at 5% level with the inf test, whereas the Yen and Euro do not reject covariance constancy at 10% level when the sup or inf tests are used.

[7] Both BIC and Phillips–Perron unit root tests are conducted on the real exchange rates of the seven base currencies from 1973:4 to 2008:12. BIC concludes that all series except the British Pound are stationary, whereas the Phillips–Perron test shows all series as stationary. Therefore, our cointegration analysis focuses on nominal exchange rates.

[8] The first term $|S_{00}|$ in $|\hat{\Sigma}(r)|$ was not computed for simplicity. Hence, AIC and BIC are both normalized to 0 for $r = 0$.

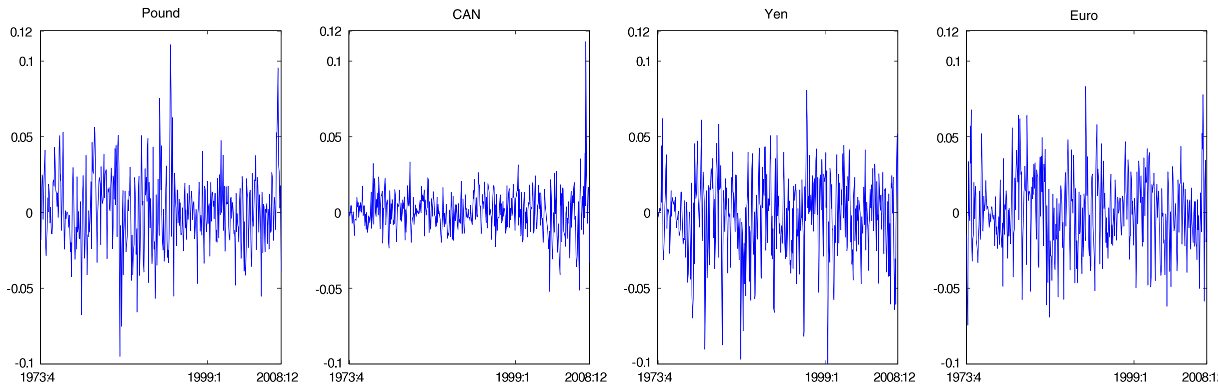

Fig. 1. Difference of log exchange rate.

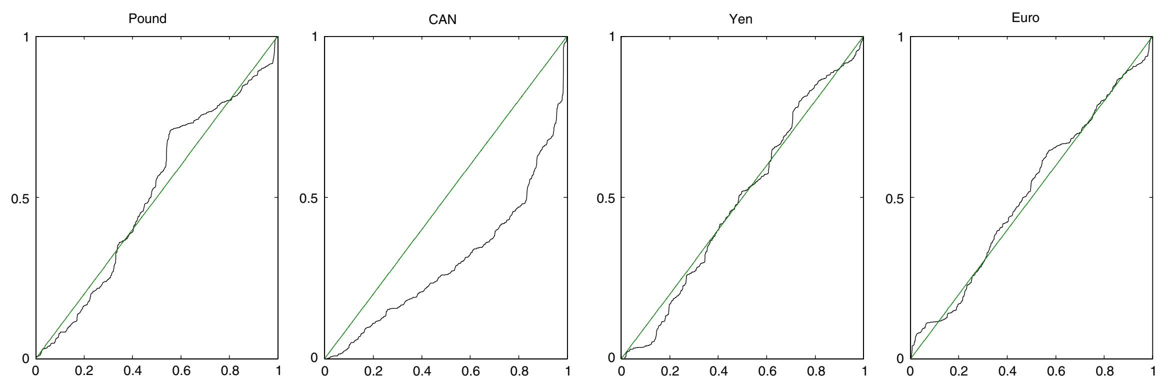

Fig. 2. Estimated variance profile.

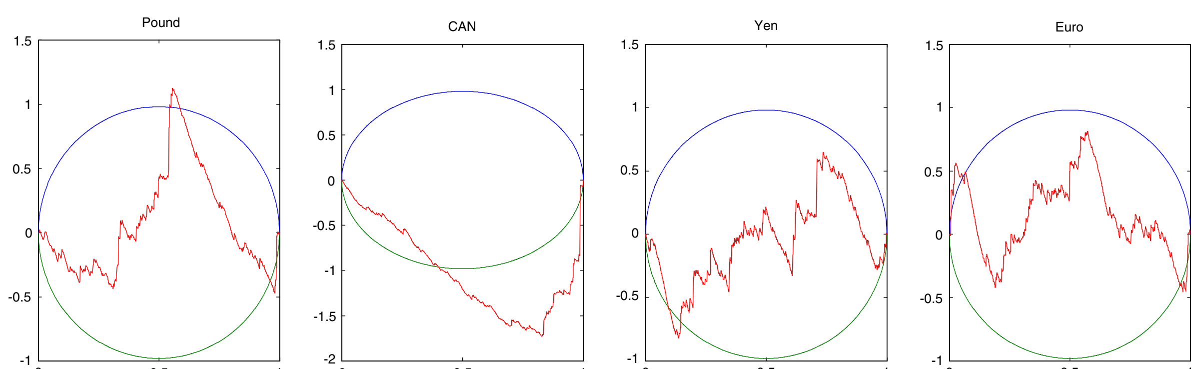

Fig. 3. 95% confidence band and sample path for the cusum of squares test.

for stochastic volatility shifts seems important for practical work, especially in financial econometric applications. Second, there is scope for using BIC to test for a shifts in variance while jointly conducting cointegrating rank estimation. Finally, models of fractional cointegration might be encompassed by using a multivariate version of the exact local Whittle procedure (Shimotsu and Phillips, 2005) to jointly estimate the fractional differencing parameters and a reduced rank coefficient matrix, by means of which cointegrating rank might be assessed as in the much simpler model (1) used here.

Appendix

We start with two preliminary results that are useful in what follows.

Lemma 3. *Under Assumptions 1–3, if* $a_t = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$ *and* $b_t = B(L)\varepsilon_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j}$ *with* $\sum_{j=0}^{\infty} j||A_j|| < \infty$ *and* $\sum_{j=0}^{\infty} j||B_j|| < \infty$*. Then*

$$n^{-1}\sum_{t=1}^{n} a_t b_{t+h}' \underset{a.s.}{\to} \sum_{j=0}^{\infty} A_j \overline{V} B_{j+h}',$$

$$n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{t-1} E(a_j b_t') \to \sum_{h=1}^{\infty}\sum_{j=0}^{\infty} A_j \overline{V} B_{j+h}'.$$

Proof of Lemma 3. Using the fact that $\varepsilon_t = V(\frac{t}{n})e_t$ and e_t is *iid* $(0, \Sigma_e)$, we have

$$n^{-1}\sum_{t=1}^{n} E(\varepsilon_t \varepsilon_t') = n^{-1}\sum_{t=1}^{n}\left(V\left(\frac{t}{n}\right)(Ee_t e_t')V\left(\frac{t}{n}\right)'\right)$$
$$\to \int_0^1 V(r)\Sigma_e V(r)' dr.$$

Since ε_t is a martingale difference sequence, we find, e.g. as in Phillips and Solo (1992), that

$$n^{-1}\sum_{t=1}^{n} a_t b_{t+h}' \to_{a.s.} \sum_{j=0}^{\infty} A_j \left(\int_0^1 V(r)\Sigma_e V(r)'dr\right) B_{j+h}' = \sum_{j=0}^{\infty} A_j \overline{V} B_{j+h}'. \tag{14}$$

Next, note that

$$\begin{aligned} n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{t-1} E(a_j b_t') &= n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} E(a_t b_{t+h}') \\ &= n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}\sum_{j=0}^{\infty} A_j V \\ &\quad \times \left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \\ &= \sum_{j=0}^{L} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} A_j V \\ &\quad \times \left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \\ &\quad + \sum_{j=L+1}^{\infty} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} A_j V \\ &\quad \times \left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}', \end{aligned} \tag{15}$$

for any integer $L > 0$ chosen so that $\frac{L}{n} + \frac{1}{L} \to 0$. For each fixed $j \le L$, we have

$$n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' = \sum_{h=1}^{n-1}\omega_{nh}\sum_{k=1}^{h} B_{j+k}', \tag{16}$$

where

$$\omega_{nh} = n^{-1}V\left(\frac{n-h-j}{n}\right)\Sigma_e V\left(\frac{n-h-j}{n}\right)'.$$

As $n \to \infty$, the sum involving ω_{nh} satisfies

$$\begin{aligned} \sum_{h=1}^{n-1}\omega_{nh} &= \sum_{h=1}^{n-1} n^{-1}V\left(\frac{n-h-j}{n}\right)\Sigma_e V\left(\frac{n-h-j}{n}\right)' \\ &= \sum_{h=1}^{n-1}\int_{\frac{h}{n}}^{\frac{h+1}{n}} V\left(\frac{n-[nr]-j}{n}\right)\Sigma_e V \\ &\quad \times \left(\frac{n-[nr]-j}{n}\right)' dr \\ &= \int_{\frac{1}{n}}^{\frac{n}{n}} V\left(\frac{n-[nr]-j}{n}\right)\Sigma_e V\left(\frac{n-[nr]-j}{n}\right)' dr \\ &\to \int_0^1 V(r)\Sigma_e V(r)'dr = \overline{V}, \end{aligned} \tag{17}$$

uniformly in j, for $j \le L$. By (16), (17), and the Toeplitz Lemma, we then have

$$n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \to \sum_{h=1}^{\infty} A_j \overline{V} B_{j+h}',$$

uniformly in j, for $j \le L$. As a result,

$$\sum_{j=0}^{L} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \to \sum_{j=0}^{\infty}\sum_{h=1}^{\infty} A_j \overline{V} B_{j+h}', \tag{18}$$

as $n \to \infty$ and $L \to \infty$.

Let C be a positive constant such that $V(r)$ is uniformly bounded above by CI_m for $r \in (-\infty, 1]$. Then

$$\left\|\sum_{j=L+1}^{\infty} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h} A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}'\right\| \le C^2\|\Sigma_e\|\sum_{j=L+1}^{\infty}\|A_j\|\sum_{h=1}^{\infty}\|B_{j+h}'\| \to 0, \tag{19}$$

as $L \to \infty$ since $\sum_{j=0}^{\infty}||A_j|| < \infty$, $\sum_{j=0}^{\infty}||B_j|| < \infty$.

It follows from (15), (18) and (19) that

$$n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{t-1} E(a_j b_t') \to \sum_{h=1}^{\infty}\sum_{j=0}^{\infty} A_j \overline{V} B_{j+h}'. \quad \square$$

Proof of Lemma 1. This is a vector generalization of Theorem 1 of Cavaliere and Taylor (2007). Using the Phillips–Solo device,

$$n^{-1/2}\sum_{s=1}^{[nr]} u_s = D(1)\sum_{s=1}^{[nr]} V\left(\frac{s}{n}\right)\frac{e_s}{\sqrt{n}} + o_p(1) \Rightarrow D(1)\int_0^r V(s)dB_e(s), \tag{20}$$

where $B_e(\cdot)$ a m-vector Brownian motion with variance Σ_e. Under Assumption 1, the $o_p(1)$ term in (20) can be verified in the same way as in Cavaliere and Taylor (2007).

By Lemma 2 of Cavaliere (2004),

$$\int_0^r V_k(s)dB(s) = \sigma_k B(\eta_k(r)),$$

for $k = 1, \ldots, m$. Because $V(\cdot)$ is diagonal, we have

$$\int_0^r V(s)dB_e(s) = \begin{pmatrix} \int_0^r V_1(s)dB_1(s) \\ \vdots \\ \int_0^r V_m(s)dB_k(s) \end{pmatrix}\Sigma_e^{1/2} = \Omega\begin{pmatrix} B_1(\eta_1(r)) \\ \vdots \\ B_k(\eta_m(r)) \end{pmatrix}\Sigma_e^{1/2}, \tag{21}$$

where $\Omega = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ and $B_k(\cdot)$, for $k = 1, \ldots, m$, are all standard Brownian motions that are independent of each other. By (20) and (21) we obtain

$$n^{-1/2}\sum_{s=1}^{[n\cdot]} u_s \Rightarrow B_V(\cdot) := D(1)\Omega B_\eta(\cdot)\Sigma_e^{1/2},$$

where $B_\eta(\cdot) = (B_1(\eta_1(\cdot)), \ldots, B_m(\eta_m(\cdot)))'$. $\square$

Proof of Lemma 2. This lemma follows the proof of Lemma 3.2 of CP by replacing $B_u(r)$ with the variance transformed Brownian motion $B_V(r)$ and by using the limit theory given in Lemma 4 below (rather than Lemma 3.1 of CP), which is of some independent interest in this context. $\square$

Proof of Theorem 1. The proof follows in the same way as the proof of Theorem 3.1 of CP and uses Lemma 2 here in place of Lemma 3.2 in CP. □

Define

$$S_{00}=n^{-1}\sum_{t=1}^{n}\Delta X_t\Delta X_t',\qquad S_{11}=n^{-1}\sum_{t=1}^{n}X_{t-1}X_{t-1}',$$
$$S_{01}=n^{-1}\sum_{t=1}^{n}\Delta X_tX_{t-1}',\quad\text{and}\quad S_{10}=n^{-1}\sum_{t=1}^{n}X_{t-1}\Delta X_t'.$$

Lemma 4. *Under Assumptions 1–3,*

$$S_{00}\to_p\Sigma_{00},\qquad \beta'S_{11}\beta\to_p\Sigma_{\beta\beta},\qquad \beta'S_{10}\to_p\Sigma_{\beta 0}$$
$$n^{-1}\beta_\perp'S_{11}\beta_\perp\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left(\int_0^1B_VB_V'\right)\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1},$$
$$\beta_\perp'S_{11}\beta\Rightarrow-(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left(\int_0^1B_VdB_V'\right)\beta(\alpha'\beta)^{-1}+\Psi_{wv},$$
$$\beta_\perp'S_{10}\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left(\int_0^1B_VdB_V'\right)\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'+\Psi_{wu}^1+\Psi_{wv}\alpha'.$$

Proof of Lemma 4. Note that $\Delta X_t=W(L)\varepsilon_t$ and $\beta'X_t=G(L)\varepsilon_t$ and the lag polynomials $W(L)$ and $G(L)$ satisfy the conditions of Lemma 3 by virtue of Assumptions 1 and 3, we have

$$S_{00}=n^{-1}\sum_{t=1}^{n}\Delta X_t\Delta X_t'\to_p\sum_{j=0}^{\infty}W_j\bar{V}W_j'=\Sigma_{00},$$
$$\beta'S_{11}\beta=n^{-1}\sum_{t=1}^{n}\beta'X_{t-1}(\beta'X_{t-1})'\to_p\sum_{j=0}^{\infty}G_j\bar{V}G_j'=\Sigma_{\beta\beta},$$
$$\beta'S_{10}=n^{-1}\sum_{t=1}^{n}\beta'X_{t-1}\Delta X_t'\to_p\sum_{j=0}^{\infty}G_j\bar{V}W_{j+1}'=\Sigma_{\beta\beta}.$$

Next we show that

$$n^{-1/2}\sum_{s=1}^{[n\cdot]}\beta'X_s\Rightarrow-(\beta'\alpha)^{-1}\beta'B_V(\cdot),$$
$$n^{-1/2}\beta_\perp'X_{[n\cdot]}\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'B_V(\cdot).\tag{22}$$

As in CP, we have the Wold representation of $\beta'X_t$

$$v_t:=\beta'X_t=\sum_{i=0}^{\infty}R^i\beta'u_{t-i}=R(L)\beta'u_t=R(L)\beta'D(L)\varepsilon_t,\tag{23}$$

and the partial sum (or generalized Granger) representation

$$X_t=C\sum_{s=1}^{t}u_s+\alpha(\beta'\alpha)^{-1}R(L)\beta'u_t+CX_0,\tag{24}$$

where $C=\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'$. In view of (24) we have

$$\begin{aligned}\beta_\perp'X_t&=\beta_\perp'C\sum_{s=1}^{t}u_s+\beta_\perp'\alpha(\beta'\alpha)^{-1}R(L)\beta'u_t+\beta_\perp'CX_0\\&=(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left\{\sum_{s=1}^{t}u_s+X_0\right\}+\beta_\perp'\alpha(\beta'\alpha)^{-1}R(L)\beta'u_t,\end{aligned}$$

so that the standardized process $n^{-1/2}\beta_\perp'X_{[n\cdot]}\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'B_V(\cdot)$. Using (23) and the fact that $R(1)=\sum_{i=0}^{\infty}R^i=(I-R)^{-1}=-(\beta'\alpha)^{-1}$, we have

$$n^{-1/2}\sum_{s=1}^{[n\cdot]}\beta'X_s\Rightarrow-(\beta'\alpha)^{-1}\beta'B_V(\cdot).\tag{25}$$

Using (22) and Lemma 3, it follows from Park and Phillips (1988) that

$$n^{-1}\beta_\perp'S_{11}\beta_\perp\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\left(\int_0^1B_VB_V'\right)\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1},$$

$$\begin{aligned}\beta_\perp'(S_{10}-S_{11}\beta\alpha')&=\beta_\perp'\left\{n^{-1}\sum_{t=1}^{n}X_{t-1}(\Delta X_t-\alpha\beta'X_{t-1})'\right\}\\&=\sum_{t=1}^{n}\frac{\beta_\perp'X_{t-1}}{\sqrt{n}}\frac{u_t}{\sqrt{n}}\\&\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1B_VdB_V'+\Psi_{wu}^1,\end{aligned}\tag{26}$$

$$\begin{aligned}\beta_\perp'S_{11}\beta&=\sum_{t=1}^{n}\frac{\beta_\perp'X_{t-1}}{\sqrt{n}}\frac{(\beta'X_{t-1})'}{\sqrt{n}}\\&\Rightarrow-(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1B_VdB_V'\beta(\alpha'\beta)^{-1}+\Psi_{wv},\end{aligned}\tag{27}$$

where

$$\begin{aligned}\Psi_{wu}^1&=\lim_{n\to\infty}n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{t-1}E((\beta_\perp'\Delta X_{t-1})u_t')\\&=\sum_{h=1}^{\infty}\sum_{j=0}^{\infty}\beta_\perp'W_j\bar{V}D_{j+h},\end{aligned}$$

$$\begin{aligned}\Psi_{wv}&=\lim_{n\to\infty}n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{t-1}E((\beta_\perp'\Delta X_{t-1})(\beta'X_{t-1})')\\&=\sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\beta_\perp'W_j\bar{V}G_{j+h}.\end{aligned}$$

Finally, using (26) and (27), we obtain

$$\begin{aligned}\beta_\perp'S_{10}&=\beta_\perp'(S_{10}-S_{11}\beta\alpha')+\beta_\perp'S_{11}\beta\alpha'\\&\Rightarrow(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp'\int_0^1B_udB_u'\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'\\&\quad+\Psi_{wu}^1+\Psi_{wv}\alpha',\end{aligned}$$

since $\beta(\alpha'\beta)^{-1}\alpha'+\alpha_\perp(\beta_\perp'\alpha_\perp)^{-1}\beta_\perp'=I$ (e.g. Johansen, 1995, p. 39). □

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest.

Akaike, H., 1977. On entropy maximization principle. In: Krishnarah, P.R. (Ed.), Applications of Statistics. North-Holland, Amsterdam.

Baillie, R.T., Bollerslev, T., 1989. Common stochastic trends in a system of exchange rates. Journal of Finance 44, 167–181.

Baillie, R.T., Bollerslev, T., 1994. Cointegration, fractional cointegration, and exchange rate dynamics. Journal of Finance 49, 737–745.

Beare, B., 2007. Robustifying unit root tests to permanent changes in innovation variance. Working paper, Yale University.

Boswijk, H.P., 2005. Adaptive testing for a unit root with nonstationary volatility. Working paper 2005/07, University of Amsterdam.

Busetti, F., Taylor, A.M.R., 2003. Testing against stochastic trend in the presence of variance shifts. Journal of Business and Economic Statistics 21, 510–531.

Cavaliere, G., 2004. Unit root tests under time-varying variance shifts. Econometric Reviews 23, 259–292.

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

Cavaliere, G., Taylor, A.M.R., 2005. Stationarity tests under time-varying variances. Econometric Theory 21, 1112–1129.
Chao, J., Phillips, P.C.B., 1999. Model selection in partially non-stationary vector autoregressive processes with reduced rank structure. Journal of Econometrics 91, 227–271.
Cheng, X, Phillips, P.C.B., 2009a. Semiparametric cointegrating rank selection. The Econometrics Journal 12, 83–104.
Cheng, X, Phillips, P.C.B., 2009b. Cointegrating Rank Selection in Models with Time-varying Variance. Cowles Foundation Discussion Paper No. 1688.
Corbae, D, Ouliaris, S., 1988. Cointegration and tests of purchasing power parity. The Review of Economics and Statistics 70, 508–511.
Diebold, F.X., Gardeazabal, J., Yilmaz, K., 1994. On cointegration and exchange rate dynamics. Journal of Finance 49, 727–735.
Hamori, S., Tokihisa, A., 1997. Testing for a unit root in the presence of a variance shift. Economics Letters 57, 245–253.
Hannan, E.J., Quinn, B.G., 1979. The determination of the order of an autoregression. Journal of the Royal Statistical Society, Series B 41, 190–195.
Hassler, U., Marmol, F., Velasco, C., 2006. Residual log-periodogram inference for long-run relationships. Journal of Econometrics 130, 165–207.
Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.
Johansen, S., 1995. Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford.
Kim, C.S., Phillips, P.C.B., 2001. Fully modified estimation of fractional cointegration models. Working Paper, Yale University.
Kim, T.H., Leybourne, S., Newbold, P., 2002. Unit root tests with a break in innovation variance. Journal of Econometrics 109, 365–387.
Loretan, M., Phillips, P.C.B., 1994. Testing covariance stationarity of heavy tailed time series: an overview of the theory with applications to several financial datasets. Journal of Empirical Finance 1, 211–248.
Nielsen, M.Ø., 2004. Optimal residual-based tests for fractional cointegration and exchange rate dynamics. Journal of Business and Economic Statistics 22, 331–345.
Nielsen, M.Ø., Shimotsu, K., 2007. Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. Journal of Econometrics 141, 574–596.
Pagan, A.R., Schwert, G.W., 1990. Testing for covariance stationarity in stock market data. Economics Letters 33, 165–170.
Park, J.Y., Phillips, P.C.B., 1988. Statistical inference in regressions with integrated processes: part 1. Econometric Theory 4, 468–497.
Phillips, P.C.B., 1996. Econometric model determination. Econometrica 64, 763–812.
Phillips, P.C.B., 2008. Unit root model selection. Journal of the Japan Statistical Society 38, 65–74.
Phillips, P.C.B., 1987. Time series regression with a unit root. Econometrica 55, 277–301.
Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. Econometrica 63, 1023–1078.
Phillips, P.C.B., McFarland, J., 1997. Forward exchange market unbiasedness: the case of the australian dollar since 1984. Journal of International Money and Finance 16, 885–907.
Phillips, P.C.B., Ploberger, W., 1996. An asymptotic theory of bayesian inference for time series. Econometrica 64, 381–413.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971–1001.
Phillips, P.C.B., Xu, K.-L., 2006. Inference in autoregression under heteroskedasticity. Journal of Time Series Analysis 27, 289–308.
Rissanen, J., 1978. Modeling by shortest data description. Automatica 14, 465–471.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Shimotsu, K., Phillips, P.C.B., 2005. Exact local whittle estimation of fractional integration. Annals of Statistics 33, 1890–1933.
Watson, M.W., 1999. Explaining the increased variability in long-term interest rates. Federal Reserve Bank Richmond Economic Quarterly 85, 71–96.
Wei, C.Z., 1992. On predictive least squares principles. Annals of Statistics 20, 1–42.
Xu, K.-L., Phillips, P.C.B., 2008. Adaptive estimation of autoregressive models with time-varying variances. Journal of Econometrics 142, 265–280.