COINTEGRATING RANK SELECTION IN MODELS WITH TIME-VARYING VARIANCE

By

Xu Cheng and Peter C. B. Phillips

January 2009

COWLES FOUNDATION DISCUSSION PAPER NO. 1688

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

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

Cointegrating Rank Selection in Models with Time-Varying Variance*

Xu Cheng
Department of Economics
Yale University

and

Peter C. B. Phillips
Cowles Foundation, Yale University
University of Auckland & Singapore Management University

October 11, 2008

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 (2008) 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.

Keywords: Cointegrating rank, Consistency, Heterogeneity, Information criteria, Model selection, Nonparametric, Time varying variances, Unit roots.

JEL classification: C22, C32

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)

*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.

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 *et al* (2002) 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 (2006) 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 (1987, 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 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, variable choice in regression, and cointegrating rank selection in parametric settings (Phillips, 1996). Cheng and Phillips (2008) show that cointegrating rank selection by suitable information criteria is consistent in a more general semiparametric framework using reduced rank regression (RRR) in a simple

VAR model with one lag. In particular, RRR may be implemented without explicitly taking into account weak dependence in the errors. The present paper strengthens the results in Cheng and Phillips (2008) 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

We consider the semiparametric ECM model

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

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

$$\begin{aligned} u_t &= D(L)\varepsilon_t = \sum_{j=0}^{\infty} D_j \varepsilon_{t-j}, \\ \varepsilon_t &= V\left(\frac{t}{n}\right) e_t, \quad e_t \sim iid\ (0, \Sigma_e), \end{aligned} \tag{2}$$

where $V(\cdot) = diag\{V_1(\cdot), \cdots, V_m(\cdot)\}$ and $V_k(\cdot)$, for $k = 1, \cdots, m$, is an unknown positive scale function. Under this specification, the innovation term ε_t has mean zero and time-varying variance $V\left(\frac{t}{n}\right)\Sigma_e V\left(\frac{t}{n}\right)$. 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 $\{\varepsilon_t\}$.

Assumption 1 below imposes conditions on the linear process u_t that facilitate the partial sum limit theory. Assumption 2 gives conditions on the innovation variance that are analogous to those used in Phillips and Xu (2006). The conditions in Assumption 3 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 $k = 1, ..., m$, 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 $\left| I_m - (I_m + \alpha\beta')L \right| = 0$ 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.*

Under (1), the time series X_t is cointegrated with cointegration matrix β of rank r_0, so there are r_0 cointegrating relations in the true model. As in Cheng and Phillips (2008), we treat (1) semiparametrically with regard to u_t and to estimate r_0 directly in (1) by information criteria. The procedure we use here is identical to that of Cheng and Phillips (2008) and is straightforward. Model (1) is estimated by conventional RRR for all values of $r = 0, 1, \cdots, m$ just as if u_t were a martingale difference, and r is chosen to optimize the corresponding information criteria as if (1) were a correctly specified parametric framework up to the order parameter r. Thus, no account is taken of the weak dependence structure and time-varying variance of u_t in the process.

Following Cheng and Phillips (2008), the information criterion used to evaluate cointegrating rank is

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

with coefficient $C_n = \log n$, $\log\log n$, or 2 corresponding to the BIC (Schwarz, 1978; Akaike, 1977; Rissanen, 1978), Hannan and Quinn (1979), and Akaike (1974) penalties, respectively, or even sample information-based versions (Wei, 1992; Phillips and Ploberger, 1996). The BIC version of (3) was given in Phillips and McFarland (1997). In (3) the degrees of freedom term $2mr - r^2$ is calculated to account for the $2mr$ elements of the matrices α and β that have to be estimated, adjusted for the r^2 restrictions that are needed to ensure structural identification of β in RRR.

The procedure is now the same as in Cheng and Phillips (2008). Only the limit theory differs because this depends on the persistent shifts in volatility. For each $r = 0, 1, \cdots, m$, we estimate the $m \times r$ matrices α and β by RRR and, for use in (3), we form the corresponding residual variance matrices

$$\widehat{\Sigma}(r) = n^{-1} \sum_{t=1}^{n} \left(\Delta X_t - \hat{\alpha}\hat{\beta}' X_{t-1}\right)\left(\Delta X_t - \hat{\alpha}\hat{\beta}' X_{t-1}\right)', \quad r = 1, ..., m$$

with $\widehat{\Sigma}(0) = n^{-1} \sum_{t=1}^{n} \Delta X_t \Delta X_t'$. Then, r is selected as $\widehat{r} = \text{argmin}_{0 \leq r \leq m} IC(r)$.

Define

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

and then (Johansen, 1995)

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

where $\widehat{\lambda}_i$, $1 \leq i \leq r$, are the r largest solutions to the determinantal equation

$$\left|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}\right| = 0. \tag{5}$$

The criterion (3) is then well determined for any given value of r.

3 Asymptotic Results

Assumption 3 ensures that the matrix $\beta'\alpha$ has full rank. 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 $\beta_\perp' \beta_\perp = I_{m-r}$. As in Cheng and Phillips (2008), we have the Wold representation of $\beta' X_t$

$$v_t := \beta' X_t = \sum_{i=0}^{\infty} R^i \beta' u_{t-i} = R(L) \beta' u_t = R(L) \beta' D(L) \varepsilon_t, \tag{6}$$

and the partial sum (or generalized Granger) representation

$$X_t = C \sum_{s=1}^{t} u_s + \alpha \left(\beta'\alpha\right)^{-1} R(L) \beta' u_t + C X_0, \tag{7}$$

where $C = \beta_\perp \left(\alpha_\perp' \beta_\perp\right)^{-1} \alpha_\perp'$.

For $k = 1, ..., m$, we define

$$\eta_k(r) := \left(\int_0^1 V_k(s)\,ds\right)^{-1} \int_0^r V_k(s)\,ds \text{ and } \sigma_k := \int_0^1 V_k(s)\,ds. \tag{8}$$

The volatility of the k'th element of ε_t is characterized by its variance profile $\eta_k(r)$, which is equal to r only when the innovation is homogeneous. The variance profile $\eta_k(r)$ is normalized by the average innovation variance σ_k so that $\eta_k(r)$ is an increasing homeomorphism on $[0,1]$ with $\eta_k(0) = 0$ and $\eta_k(1) = 1$.

Lemma 1 *Under Assumptions 1-3,*

(a)

$$n^{-1/2}\sum_{s=1}^{[n\cdot]} u_s \Rightarrow D(1)\,B_\varepsilon(\cdot)\,, \ \textit{where } B_\varepsilon(r) = \int_0^r V(s)\,dB_e(s)\,,$$

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

(b)

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

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

(c)

$$\begin{aligned} n^{-1/2}\sum_{s=1}^{[n\cdot]} v_s &\Rightarrow -\left(\beta'\alpha\right)^{-1}\beta' B_V(\cdot)\,, \\ n^{-1/2}\beta_\perp' X_{[n\cdot]} &\Rightarrow \left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp' B_V(\cdot)\,. \end{aligned}$$

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

Let

$$v_t = G(L)\varepsilon_t \text{ and } \Delta X_t = \alpha v_{t-1} + u_t = W(L)\varepsilon_t, \tag{9}$$

where $G(L) = R(L)\,\beta' D(L)$ by (6) and $W(L) = \alpha L G(L) + D(L)$. Define the average variance of ε_t by

$$\overline{V} := \int_0^1 V(r)\,\Sigma_e V(r)'\,dr. \tag{10}$$

The following results provide some asymptotic limits that are useful in deriving the asymptotic properties of $\widehat{\Sigma}(r)$, extending a corresponding result in Cheng and Phillips (2008).

Lemma 2 *Under Assumptions 1-3,*

$$S_{00} \to_p \Sigma_{00},\ \beta' S_{11}\beta \to_p \Sigma_{\beta\beta},\ \beta' S_{10} \to_p\ \Sigma_{\beta 0}$$

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

$$\beta_\perp' S_{11}\beta \Rightarrow -\left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\left(\int_0^1 B_V dB_V'\right)\beta(\alpha'\beta)^{-1} + \Psi_{wv},$$

$$\beta_\perp' S_{10} \Rightarrow \left(\alpha_\perp'\beta_\perp\right)^{-1}\alpha_\perp'\left(\int_0^1 B_V dB_V'\right)\alpha_\perp\left(\beta_\perp'\alpha_\perp\right)^{-1}\beta_\perp' + \Psi_{wu}^1 + \Psi_{wv}\alpha',$$

where

$$\Sigma_{00} = \sum_{j=0}^{\infty} W_j\overline{V}W_{j+h}',\ \Sigma_{\beta\beta} = \sum_{j=0}^{\infty} G_j\overline{V}G_{j+h}',\ \Sigma_{\beta 0} = \sum_{j=0}^{\infty} G_j\overline{V}W_{j+h}', \tag{11}$$

$$\Psi_{wu}^1 = \sum_{h=1}^{\infty}\sum_{j=0}^{\infty}\beta_\perp' W_j\overline{V}D_{j+h},\ \Psi_{wv} = \sum_{h=0}^{\infty}\sum_{j=0}^{\infty}\beta_\perp' W_j\overline{V}G_{j+h}, \tag{12}$$

and $w_t = \beta_\perp'\Delta X_t = \beta_\perp' W(L)\varepsilon_t$.

Remarks: When the innovation ε_t has time-varying variances, the asymptotic limit associated with the non-stationary process $\beta_\perp' X_t$ involves the variance transformed Brownian motion $B_V(\cdot)$. 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_{wu}^1 and Ψ_{wv} are all simplified to moments of ΔX_t and v_t, both of which are now stationary. Those results under homogeneous errors were given in Cheng and Phillips (2008).

Define

$$\widetilde{\alpha} = \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1} \tag{13}$$

and let $\widetilde{\alpha}_\perp$ be an $m\times(m-r)$ orthogonal complement to $\widetilde{\alpha}$ such that $[\widetilde{\alpha}, \widetilde{\alpha}_\perp]$ is nonsingular. The following reproduces Lemma 2 in Cheng and Phillips (2008) under shifting variances.

Lemma 3 *Under Assumptions 1-3, when the true cointegration rank is r_0, the r_0 largest solutions to (5), denoted by $\widehat{\lambda}_i$ with $1 \le i \le r_0$, converge to the roots of*

$$\left|\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}\right| = 0. \tag{14}$$

The remaining $m - r_0$ roots, denoted by $\widehat{\lambda}_i$ with $r_0 + 1 \le i \le m$, decrease to zero at the rate n^{-1} and $\{n\widehat{\lambda}_i : i = r_0 + 1, ..., m\}$ converge weakly to the roots of

$$\left|\rho\int_0^1 G_u G_u' - \left(\int_0^1 G_u dG_u'\beta_\perp' + \Psi\right)\widetilde{\alpha}_\perp\left(\widetilde{\alpha}_\perp'\Sigma_{00}\widetilde{\alpha}_\perp\right)^{-1}\widetilde{\alpha}_\perp'\left(\beta_\perp\int_0^1 dG_u G_u' + \Psi'\right)\right| = 0, \tag{15}$$

where $G_u(r) = (\alpha_\perp' \beta_\perp)^{-1} \alpha_\perp' B_V(r)$ *is* $m - r_0$ *dimensional variance transformed Brownian motion and* $\Psi = \Psi_{wu}^1 + \Psi_{wv}\alpha'$.

Comparing Theorem 3 with the results under homogeneous or martingale difference errors, we see that in all cases the r_0 largest roots of (5) are all positive in the limit and the $m - r_0$ smallest roots converge to 0 at the rate n^{-1}. However, under the present set-up, the determinantal equation (15) involves a variance transformed Brownian motion G_u, which reduces to Brownian motion when the innovation is homogeneous, as in Cheng and Phillips (2008).

Allowing for weak dependence in the errors, equation (15) involves the one sided long run variance matrix Ψ. A general form of the one sided long run variance under weakly dependent heterogeneously distributed errors was first given in Phillips and Park (1988). Under Assumption 2, the components of $\Psi = \Psi_{wu}^1 + \Psi_{wv}\alpha'$ can be expressed as in (12) using the average innovation variance $\overline{V}$ by means of Lemma 4 in the Appendix.

The main result now extends the corresponding theorem Cheng and Phillips (2008) to allow for shifting variances.

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*

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

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

and

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

where $\xi_{r_0+1}, ..., \xi_m$ *are the ordered roots of the limiting determinantal equation* (15).

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.

AIC is inconsistent, asymptotically never underestimates cointegrating rank, and favors more liberally parametrized 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 homogenous 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 Cheng and Phillips (2008). 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:

$$A \ : \ \alpha'\beta = \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.4 \end{pmatrix}, \ \text{ with stationary roots } \lambda_i \left[I + \beta'\alpha\right] = \{0.7, 0.4\};$$

$$B \ : \ \alpha'\beta = \begin{pmatrix} -0.5 & 0.1 \\ 0.2 & -0.15 \end{pmatrix}, \ \text{ with stationary roots } \lambda_i \left[I + \beta'\alpha\right] = \{0.9, 0.45\}.$$

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

$$u_t = Au_{t-1} + \varepsilon_t, \ u_t = \varepsilon_t + B\varepsilon_{t-1}, \ u_t = Au_{t-1} + \varepsilon_t + B\varepsilon_{t-1}, \tag{16}$$

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

$$\varepsilon_t = V\left(\frac{t}{n}\right) e_t \text{ and } e_t = iid \ N\left(0, \Sigma_\varepsilon\right), \tag{17}$$

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 $r \in (-\infty, 1]$, the $m \times m$ diagonal variance matrix $V(r) = g(r)^2 I_m$, 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:

$$\begin{aligned} &1.\ g(r)^2 = \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right) 1_{\{r \geq \tau\}}, \quad r \in [0,1], \\ &2.\ g(r)^2 = \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right) 1_{\{\tau \leq r < 1-\tau\}}, \quad r \in [0,1], \ \tau \in [0, 1/2], \\ &3.\ g(r)^2 = \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right) r^m, \quad r \in [0,1]. \end{aligned} \tag{18}$$

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 setup 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[1] 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 (16). 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

[1] It is shown in Cheng and Phillips (2008) 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 $C_n = \log\log\log n$.

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 of 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

			$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
τ	δ	$\widehat{r}$	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

			$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
τ	δ	$\widehat{r}$	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

			$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
m	δ	$\widehat{r}$	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 (18) . To save space, we only report the results under model 1, as shown in Table 4. Comparing Tables 4 and 1, we find that BIC is generally more reliable when the errors have low temporal dependence. Similar results were found for models 2 and 3.

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

			$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
τ	δ	$\widehat{r}$	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

The results for design B, where the stationary roots of the system are closer to unity, are shown in Table 5. Just as in Cheng and Phillips (2008), 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.

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

			$n = 400$						$n = 100$					
			$r_0 = 0$		$r_0 = 1$		$r_0 = 2$		$r_0 = 0$		$r_0 = 1$		$r_0 = 2$	
τ	δ	$\widehat{r}$	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

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.

Our focus in this application is to apply semiparametric rank selection methods to investigate possible cointegrating relations among exchange rates under both floating exchange rate regimes (post 1973) and fixed exchange rate regimes (under the Bretton Woods agreement of 1946-1973). It is now a well-established stylized fact that many macro-economic and financial variables, including exchange rates, are characterized by breaks in volatility. So 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 data set concentrates on the same exchange rates as those in the literature cited above. The data comprise 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 January 1974 through December 2001. Our data set, taken from the DRI Economics Database (previously Citibase), is also monthly averages of noon buying rates and runs from November 1967 to December 1998[2]. The data set can be divided into two subperiods: the first period, from November 1967 to December 1973, corresponds to the fixed exchange rate regime by the Bretton Woods agreement (1946-1973); and the second period, from January 1974 to December 1998, corresponds to the floating exchange rate regime before the introduction of the Euro. Compared with earlier applications, our data for the floating exchange rate period covers a long time span of 25 years with 300 observations. We do not include observations after December 1998, as in Nielsen and Shimotsu (2007), because since then the exchange rates between some major European currencies have been fixed in relation to the Euro. The log exchange rate data series are plotted in Figure 1.

The time-varying behavior of the exchange rate volatilities is well characterized

[2]Our data, taken from the current version of the same source as that in Kim and Phillips (2001), starts from November 1967.

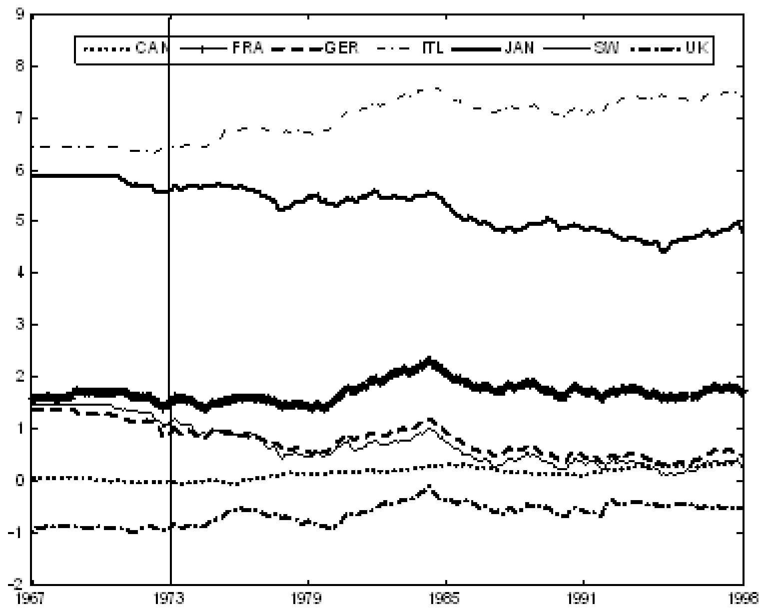

Figure 1: Log exchange rates

by its variance profile $\eta_k(r)$, for $k = 1, \cdots, m$, which is increasing from 0 to 1 and only equal to r under homogeneous errors. We first estimate the variance profile of each exchange rate series using the method of Cavaliere and Taylor (2007). Let $\{\widehat{u}_t\}$ denote the residuals from the linear regression of $\widehat{X}_t$ on $\widehat{X}_{t-1}$, where $\widehat{X}_t$ is the residual of X_t after detrending. Detrending X_t is necessary when we include an intercept in (1). The estimator of the variance profile, which is the sample analogue of (8) linearly interpolated between the observed sample data, can be written as

$$\widehat{\eta}_k(r) = \frac{\sum_{t=1}^{[nr]} \widehat{u}_t^2 - (nr - [nr])\, \widehat{u}_{[nr]+1}^2}{\sum_{t=1}^{n} \widehat{u}_t^2}. \tag{19}$$

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 Figure 2. The first two rows are the estimated variance profiles for each currency (to the US Dollar) in the post-1973 floating exchange rate period, while the last two rows are the corresponding variance profiles estimated by data within the fixed exchange rate period. The 45^{o} line corresponds to the variance profile for homogeneous errors. During the relatively long time span after 1973, we see that most exchange rates did not experience sharp changes in volatility, although multiple shifts and smooth transitions do exist in most series. Specifically, in this period, the Canadian Dollar has the smoothest volatility profile, followed by

French Franc, Deutsche mark and Swiss Franc, whose volatility generally exhibits a smooth increase at the beginning of the period and a smooth decrease at the end of the period. Compared with these currencies, the Lira had sharper positive shifts at both the beginning and the end of the period, each followed by an immediate sharp negative shift, and the British Pound has an abrupt positive shift near the end of the period, which is also followed by a negative shift. The variance profile of the Yen exhibits an positive shift in the beginning and several small shifts in the middle. These profiles indicate that the major European currencies are more closely related to each other and the Yen and the Canadian Dollar are relatively independent. As we can see from the last two rows of Figure 2, the volatility shifts under the fixed exchange rate regime is much steeper, partially due to the relatively short time span. At the end of the fixed exchange rate period, all currencies except the Canadian Dollar had steep increases in volatility.

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 $I(0)$. The findings confirm earlier conclusions that nominal exchange rates are well characterized as $I(1)$ processes (c.f., Corbae and Ouliaris, 1988, and Baillie and Bollerslev, 1989). Table 6 and 7[3] report results for AIC and BIC for each currency under both flexible and fixed exchange rate regimes. Both the theory and the simulation results predict that AIC generally overestimates order, which biases results to stationarity. BIC, which is more reliable, shows almost all series to be I(1) processes. The only exception is the British Pound in the fixed exchange rate period. However, from the simulation findings in the last section, this outcome for the British Pound may well be due to overestimation resulting from the huge volatility jump of the Pound at the beginning of the period.

Next, cointegrating rank among the seven exchange rates is estimated by AIC and BIC under (1). The method allows for both weak dependence and variance heterogeneity as detected in Figure 2. The estimation results are presented in Table 8 and 9. Under the floating exchange rate regime, AIC finds 4 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. Our result is consistent with that obtained using the Johansen trace test, where the optimal number of lags is selected with information criterion (Diebold, *et al,* 1994). Compared with Johansen's method, our procedure do 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.

Table 9 shows that cointegrating rank is estimated as 6 by AIC and 1 by BIC under the fixed exchange rate regime. The difference in these outcomes is substantial, but we note that: (i) simulations show that AIC has a strong tendency to overestimate

[3] AIC(0) and BIC(0) in table 1-4 are normalized to 0 for computational convenience, but this normalization does not affect estimation results.

cointegrating rank, whereas BIC shows only a small tendency to underestimate rank; and (ii) the empirical results show that the BIC estimate is more sharply determined than AIC. Taking the more reliable result given by BIC, we conclude that under the fixed exchanged rate regime different currencies were tied to one equilibrium path in the long run and that deviations from this long run path were temporary. This result is compatible with the nature of the fixed exchange rate regime, whereby under the Bretton Woods agreement, exchange rates were tied to each other, allowing some adjustments only under special circumstances. Thus, empirical confirmation of some long run equilibrium relationship among the exchange rates is to be expected during the Bretton Woods era.

Table 6

Unit root test for individual series under floating exchange rate regime

		CAN	FRA	GER	ITL	JAN	SW	UK
AIC	$r = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$r = 1$	0.0056	0.0021	-0.0015	-0.0050	0.0031	-0.0123	-0.0047
BIC	$r = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$r = 1$	0.0180	0.0145	0.0109	0.0074	0.0154	0.0001	0.0077
AIC	$\widehat{r}$	0	0	1	1	0	1	1
BIC	$\widehat{r}$	0	0	0	0	0	0	0

Table 7

Unit root test for individual series under fixed exchange rate regime

		CAN	FRA	GER	ITL	JAN	SW	UK
AIC	$r = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$r = 1$	0.0193	0.0203	0.0263	0.0052	0.0270	0.0247	-0.1223
BIC	$r = 0$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$r = 1$	0.0507	0.0517	0.0577	0.0366	0.0584	0.0561	-0.0910
AIC	$\widehat{r}$	0	0	0	0	0	0	1
BIC	$\widehat{r}$	0	0	0	0	0	0	1

Table 8

Cointegration rank estimation under floating exchange rate regime

r	0	1	2	3	4	5	6	7	$\widehat{r}$
AIC	0.00	-0.06	-0.09	-0.10	-0.11	-0.09	-0.09	-0.08	4
BIC	0.00	0.10	0.21	0.30	0.39	0.46	0.51	0.52	0

Table 9

Cointegration rank estimation under fixed exchange rate regime

r	0	1	2	3	4	5	6	7	$\widehat{r}$
AIC	0.00	-0.67	-0.87	-1.07	-1.14	-1.18	-1.22	-1.19	6
BIC	0.00	-0.26	-0.12	-0.04	0.12	0.23	0.29	0.35	1

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 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 paramaters and a reduced rank coefficient matrix, by means of which cointegrating rank might be assessed as in the much simpler model (1) used here.

7 Appendix

Lemma 4 *Under Assumption 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*

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

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

Proof of Lemma 4: Using the fact that $\varepsilon_t = V\left(\frac{t}{n}\right) e_t$ and e_t is $iid\ (0, \Sigma_e)$, we have

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

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

$$n^{-1}\sum_{t=1}^{n} a_t b_{t+h}' \rightarrow_{a.s.} \sum_{j=0}^{\infty} A_j \left(\int_0^1 V(r)\,\Sigma_e V(r)'\,dr\right) B_{j+h}' = \sum_{j=0}^{\infty} A_j \overline{V} B_{j+h}'. \tag{20}$$

Next, note that

$$
\begin{aligned}
n^{-1}\sum_{t=2}^{n}\sum_{j=1}^{t-1}E(a_j b_t') &= n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}E(a_t b_{t+h}') \\
&= n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}\sum_{j=0}^{\infty}A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \\
&= \sum_{j=0}^{L} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \\
&+ \sum_{j=L+1}^{\infty} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}', \qquad (21)
\end{aligned}
$$

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

$$
n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' = \sum_{h=1}^{n-1}\omega_{nh}\sum_{k=1}^{h}B_{j+k}', \qquad (22)
$$

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

$$
\begin{aligned}
\sum_{h=1}^{n-1}\omega_{nh} &= \sum_{h=1}^{n-1}n^{-1}V\left(\frac{n-h-j}{n}\right)\Sigma_e V\left(\frac{n-h-j}{n}\right)' \\
&= \sum_{h=1}^{n-1}\int_{\frac{h}{n}}^{\frac{h+1}{n}} V\left(\frac{n-[nr]-j}{n}\right)\Sigma_e V\left(\frac{n-[nr]-j}{n}\right)' dr \\
&= \int_{\frac{1}{n}}^{\frac{n}{n}} V\left(\frac{n-[nr]-j}{n}\right)\Sigma_e V\left(\frac{n-[nr]-j}{n}\right)' dr \\
&\to \int_0^1 V(r)\,\Sigma_e V(r)'\,dr = \overline{V}, \qquad (23)
\end{aligned}
$$

uniformly in j, for $j \le L$. By (22), (23) , 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,

$$
\sum_{j=0}^{L} n^{-1}\sum_{h=1}^{n-1}\sum_{t=1}^{n-h}A_j V\left(\frac{t-j}{n}\right)\Sigma_e V\left(\frac{t-j}{n}\right)' B_{j+h}' \to \sum_{j=0}^{\infty}\sum_{h=1}^{\infty}A_j\overline{V}B_{j+h}', \qquad (24)
$$

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

$$\left\| \sum_{j=L+1}^{\infty} n^{-1} \sum_{h=1}^{n-1} \sum_{t=1}^{n-h} A_j V\left(\frac{t-j}{n}\right) \Sigma_e V\left(\frac{t-j}{n}\right)' B'_{j+h} \right\|$$
$$\leq \quad C^2 \left\|\Sigma_e\right\| \sum_{j=L+1}^{\infty} \left\|A_j\right\| \sum_{h=1}^{\infty} \left\|B'_{j+h}\right\| \to 0, \tag{25}$$

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

It follows from (21), (24) and (25) that

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

□.

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

$$n^{-1/2} \sum_{s=1}^{[nr]} u_s = D\left(1\right) \sum_{s=1}^{[nr]} V\left(\frac{s}{n}\right) \frac{e_s}{\sqrt{n}} + o_p\left(1\right) \Rightarrow D\left(1\right) \int_0^r V\left(s\right) dB_e\left(s\right), \tag{26}$$

where $B_e(\cdot)$ a m-vector Brownian motion with variance Σ_e. Under Assumption 1, the $o_p(1)$ term in (26) can be verified in the same way as in Cavaliere and Taylor (2007).

By Lemma 2 of Cavaliere (2004),

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

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

$$\int_0^r V\left(s\right) dB_e\left(s\right) = \begin{pmatrix} \int_0^r V_1\left(s\right) dB_1\left(s\right) \\ \vdots \\ \int_0^r V_m\left(s\right) dB_k\left(s\right) \end{pmatrix} \Sigma_e^{1/2} = \Omega \begin{pmatrix} B_1\left(\eta_1\left(r\right)\right) \\ \vdots \\ B_k\left(\eta_m\left(r\right)\right) \end{pmatrix} \Sigma_e^{1/2}, \tag{27}$$

where $\Omega = diag\left(\sigma_1, ..., \sigma_m\right)$ and $B_k(\cdot)$, for $k = 1, \cdots, m$, are all standard Brownian motions that are independent of each other. By (26) and (27) we obtain

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

where $B_\eta\left(\cdot\right) = \left(B_1\left(\eta_1\left(\cdot\right)\right), \cdots, B_m\left(\eta_m\left(\cdot\right)\right)\right)'$.

In view of (7) we have

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

so that the standardized process $n^{-1/2}\beta_\perp' X_{[n\cdot]} \Rightarrow \left(\alpha_\perp'\beta_\perp\right)^{-1} \alpha_\perp' B_V\left(\cdot\right)$. Using (6) and the fact that $R\left(1\right) = \sum_{i=0}^{\infty} R^i = \left(I - R\right)^{-1} = -\left(\beta'\alpha\right)^{-1}$, we have

$$n^{-1/2} \sum_{s=1}^{[n\cdot]} \beta' X_s \Rightarrow -\left(\beta'\alpha\right)^{-1} \beta' B_V\left(\cdot\right). \tag{28}$$

□.

Proof of Lemma 2: Writing $\Delta X_t = W\left(L\right)\varepsilon_t$ and $\beta' X_t = G(L)\varepsilon_t$ as in (9) and noting that the lag polynomials $W\left(L\right)$ and $G(L)$ satisfy the conditions of Lemma 4 by virtue of Assumptions 1 and 3, we have

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

Using Lemma 1 and Lemma 4, it follows from Park and Phillips (1988) that

$$\begin{aligned}
n^{-1}\beta_\perp' S_{11} \beta_\perp &\Rightarrow \left(\alpha_\perp'\beta_\perp\right)^{-1} \alpha_\perp' \left(\int_0^1 B_V B_V' \right) \alpha_\perp \left(\beta_\perp'\alpha_\perp\right)^{-1}, \\
\beta_\perp'(S_{10} - S_{11}\beta\alpha') &= \beta_\perp' \left\{ n^{-1} \sum_{t=1}^{n} X_{t-1} \left(\Delta X_t - \alpha\beta' X_{t-1}\right)' \right\} \\
&= \sum_{t=1}^{n} \frac{\beta_\perp' X_{t-1}}{\sqrt{n}} \frac{u_t}{\sqrt{n}} \Rightarrow \left(\alpha_\perp'\beta_\perp\right)^{-1} \alpha_\perp' \int_0^1 B_V dB_V' + \Psi_{wu}^1,
\end{aligned} \tag{29}$$

$$\beta_\perp' S_{11} \beta = \sum_{t=1}^{n} \frac{\beta_\perp' X_{t-1}}{\sqrt{n}} \frac{\left(\beta' X_{t-1}\right)'}{\sqrt{n}} \Rightarrow -\left(\alpha_\perp'\beta_\perp\right)^{-1} \alpha_\perp' \int_0^1 B_V dB_V' \beta (\alpha'\beta)^{-1} + \Psi_{wv}, \tag{30}$$

where

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

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

Finally, using (29) and (30), we obtain

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

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

Proof of Lemma 3: This Lemma follows the proof of Lemma 2 of Cheng and Phillips (2008) by replacing $B_u(r)$ with the variance transformed Brownian motion $B_V(r)$. □

Proof of Theorem 1: The proof follows in the same way as the proof of Theorem 1 of Cheng and Phillips (2008). □

8 References

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

Akaike, H. (1977). "On entropy maximization principle." In P. R. Krishnarah (ed.), *Applications of Statistics*. Amsterdam: North–Holland.

Baillie, R. T. and Bollerslev, T. (1989). "Common stochastic trends in a system of exchange rates," *Journal of Finance*, 44, 167-181.

Baillie, R. T. and 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. and A. M. R. Taylor (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. and A.M. R. Taylor (2005). "Stationarity tests under time-varying variances", *Econometric Theory*, 21, pp. 1112-1129.

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

Chao, J. and P. C. B. Phillips (1999). "Model selection in partially non-stationary vector autoregressive processes with reduced rank structure," *Journal of Econometrics*, Vol. 91, 227-271.

Cheng, X and P. C. B. Phillips (2008). "Semiparametric Cointegrating Rank Selection," *The Econometrics Journal*, forthcoming.

Corbae, D and S. Ouliaris (1988), "Cointegration and Tests of Purchasing Power Parity", *The Review of Economics and Statistics* 70, 508–511.

Diebold, F. X., Gardeazabal, J. and Yilmaz, K. (1994). "On cointegration and exchange rate dynamics," *Journal of Finance*, 49, 727-735.

Hamori S. and A. Tokihasa (1997). "Testing for a unit root in the presence of a variance shift." *Economics Letters*, 57, 245-253.

Hannan, E. J. and B. G. Quinn (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.

Kim, C. S. and Phillips, P. C. B. (2001). "Fully modified estimation of fractional cointegration models." Working paper, Yale University.

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: Oxford University Press.

Loretan, M. and P. C. B. Phillips (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,* 211-48.

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. and G. W. Schwert (1990a). "Testing for covariance stationarity in stock market data," *Economics Letters* 33, 165–170.

Phillips, P. C. B. (1991). "Optimal Inference in Cointegrated Systems," *Econometrica* 59, 283–306.

Phillips, P. C. B. (1995). "Fully modified least squares and vector autoregression," *Econometrica* 63, 1023-1078.

Phillips, P. C. B. (1996). "Econometric Model Determination," *Econometrica*, 64, 763-812.

Phillips, P. C. B. (1998). "New Tools for Understanding Spurious Regressions," *Econometrica*, 66, 1299-1326.

Phillips, P. C. B. (2005). "Challenges of Trending Time Series Econometrics, *Mathematics and Computers in Simulation*," 68, 401-416.

Phillips, P. C. B. (2008). "Unit root model selection," *Journal of the Japan Statistical Society,* 38, 65-74.

Phillips, P. C. B. and J. McFarland (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. and K.-L. Xu (2006), "Inference in autoregression under heteroskedasticity," *Journal of Time Series Analysis*, 27, 289-308.

Phillips P. C. B. and W. Ploberger (1996). "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica,* 64, 381-413.

Phillips, P. C. B. and V. Solo (1992). "Asymptotics for Linear Processes," *Annals of Statistics* 20, 971–1001.

Pötscher, B. M. (1989). "Model selection under nonstationarity: Autoregressive models and stochastic linear regression models," *Annals of Statistics* 17, 1257–1274.

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. and P. C. B. Phillips (2005). "Exact Local Whittle Estimation of Fractional Integration", *Annals of Statistics*, 33, 1890-1933.

Tsay, R. S. (1984). "Order selection in nonstationary autoregressive models," *Annals of Statistics*, 12, 1425-1433.

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. and Phillips, P. C. B. (2008). "Adaptive estimation of autoregressive models with time-varying variances," *Journal of Econometrics*, 142, 265-280.

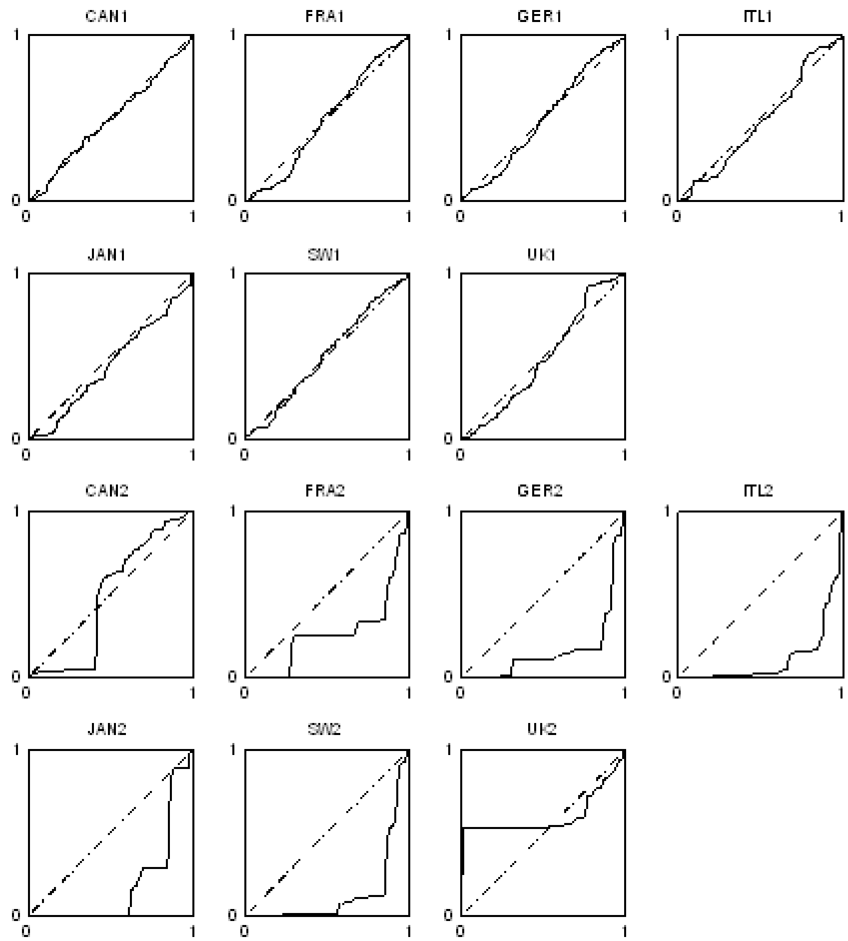

Figure 2: Estimated variance profiles of exchange rates over two periods (1: 1974-1998; and 2: 1967-1973).
CAN: Canadian Dollar; FRA: French Franc; GER: Deutsche Mark; ITL: Italian Lira; JAN: Japanese Yen; SW: Swiss Franc; UK: British Pound.