Consistent cotrending rank selection when both stochastic and nonlinear deterministic trends are present*

Zheng-Feng Guo† and Mototsugu Shintani‡

This version: February 2011

Abstract

This paper proposes a model-free cotrending rank selection procedure based on the eigenstructure of a multivariate version of the von Neumann ratio, in the presence of both stochastic and nonlinear deterministic trends. Our selection criteria are easily implemented, and the consistency of the rank estimator is established under very general conditions. Simulation results suggest good finite sample properties of the new rank selection criteria. The proposed method is then illustrated through an application of the Japanese money demand function allowing for the cotrending relationship among money, income and interest rates.

Key words: *Cointegrating rank, Smooth transition trend model, Trend breaks, Von Neumann ratio.*

*We thank Walt Enders, Yanqin Fan, Junsoo Lee, Tong Li, Ron Masulis and the seminar and conference participants at the University of Alabama-Tuscaloosa, University of Texas at Dallas, Kyoto University, Vanderbilt University, 20th Annual Meetings of the Midwest Econometrics Group and the 2010 NBER-NSF Time Series Conference for helpful comments and discussion.

†International Monetary Fund, Washington, D.C. 20431, USA. Email: `zguo@imf.org`

‡Corresponding author, Department of Economics, Vanderbilt University, VU Station B #351819, Nashville, TN 37235-1819. Email: `mototsugu.shintani@vanderbilt.edu`

1 Introduction

For decades, one of the most important issues in the analysis of macroeconomic time series has been how to incorporate a trend. Two popular approaches that have often been employed in the literature are (i) to consider a stochastic trend, with or without a linear deterministic trend, such as the one suggested in Nelson and Plosser (1982), and (ii) to consider a nonlinear deterministic trend such as the one with trend breaks considered in Perron (1989, 1997). Cointegration, introduced by Engle and Granger (1987), is a useful concept in understanding the nature of comovement among variables based on the first approach. In cointegration analysis, the cointegrating rank, defined as the number of linearly independent cointegrating vectors, provides valuable information regarding the trending structure of a multivariate system with stochastic trends. Several model-free consistent cointegrating rank selection procedures have been developed in the literature. Analogous to cointegration analysis is the analysis of comovement based on the second approach, namely, the nonlinear deterministic trend. The cotrend analyses of Bierens (2000), Hatanaka (2000) and Hatanaka and Yamada (2003) lie along this line of research. However, a consistent selection procedure of the cotrending rank, defined similarly as the cointegrating rank with a stochastic trend replaced by a nonlinear deterministic trend, has not yet been developed.

This paper proposes a model-free consistent cotrending rank selection procedure when both stochastic and nonlinear deterministic trends are present in a multivariate system. Consistency here refers to the property that the probability of selecting the wrong cotrending rank approaches zero as sample size tends to infinity. Our procedure selects the cotrending rank by minimizing the von Neumann criterion, similar to the one used by Shintani (2001) and Harris and Poskitt (2004) in their analyses of cointegration. This approach exploits the fact that identification of cotrending rank can be interpreted as identification among three groups of eigenvalues of the generalized von Neumann ratio. Using this property of the von Neumann criterion, we propose two types of cotrending rank selection procedures that are (i) invariant to linear transformations of the data; (ii) robust to model misspecification; and (iii) valid not only with a break in the trend but also with a broader class of nonlinear trend functions. The simulation results also suggest that our cotrending rank selection procedures perform well in finite samples.

Our analysis is closely related to that of Harris and Poskitt (2004) and Cheng and Phillips (2009), who propose consistent cointegrating rank procedures that do not require a parametric vector autoregressive model of cointegration such as the one in Johansen (1991). While we provide some examples of nonlinear trend functions, including trend breaks and smooth transition trend models, our cotrending rank selection procedure does not require the parametric specification of the trend function, or the parametric specification of serial dependence structure. Thus, our approach generalizes the results of Harris and Poskitt (2004) and Cheng and Phillips (2009) in the sense that it allows both common stochastic trends and common deterministic trends. Consequently, we can also use our procedure to determine the cointegrating rank in the absence of nonlinear deterministic trends. To illustrate this feature, we include both cointegrated and cotrended cases in our simulation analysis.

As emphasized in Stock and Watson (1988), the cointegrated system can be interpreted as a factor model with a stochastic trend as a common factor. Thus, determining the cointegrating rank is identical to determining the number of common stochastic trends because the latter is the difference between the dimension of the system (number of variables) and the cointegrating rank.[1] In the presence of both stochastic and nonlinear deterministic trends, however, the number of common nonlinear deterministic trends does not correspond to the difference between the dimension and the cotrending rank. Because the number of common deterministic trends also contains valuable information about the trending structure, we introduce the notion of weak cotrending rank, so that the difference between the dimension and the weak cotrending rank becomes the number of common deterministic trends.

Our two alternative definitions of cotrend are a natural consequence of the notion of a common

[1] The PANIC method proposed by Bai and Ng (2004) utilizes a consistent selection of the number of common stochastic trends in a very large dynamic factor system based on information criteria. See also Bai and Ng (2002) for the case of consistent selection of the number of stationary common factors.

feature introduced in Engle and Kozicki (1993). They define the common feature as "a feature that is present in each of a group of series but there exists a non-zero linear combination of the series that does not have the feature." When such a feature is a *broad class* of trends, namely, a mixture of both stochastic and deterministic trends, the definition of cotrend requires a linear combination that eliminates both types of trends at the same time. In contrast, when such a feature is the *dominant* trend, namely, the deterministic trend alone, a linear combination should eliminate the deterministic trend but not necessarily the stochastic trend. Since the latter type of cotrend nests the former type, we distinguish between the two by referring to the latter type as a weaker version of the cotrending relationship. Our procedure is designed to determine both the cotrending rank and weak cotrending rank.

The remainder of this paper is organized as follows. Section 2 introduces some key concepts in the system of common stochastic and deterministic trends. The main theoretical results are provided in section 3. Section 4 reports Monte Carlo simulation results to show the finite sample performance of our procedures. In section 5, we apply our procedures to the Japanese money demand function. Section 6 concludes, and the technical proofs are presented in the Appendix.

2 Motivation

2.1 Cotrending ranks

Our cotrend analysis begins with an assumption that all the variables contain deterministic trends. This presumption is similar to the case of traditional cointegration analysis, which requires all the variables to follow I(1) processes so that at least one stochastic trend is present in each variable of interest. The following simple bivariate examples illustrate the motivation of our cotrend analysis. In the presence of deterministic trends, a pair of variables, $y_t = (y_{1t}, y_{2t})'$, can be decomposed as

$$\begin{aligned} y_{1t} &= d_{1t} + s_{1t}, \\ y_{2t} &= d_{2t} + s_{2t}, \end{aligned} \tag{1}$$

where $d_t = (d_{1t}, d_{2t})'$ represents a deterministic trend component and $s_t = (s_{1t}, s_{2t})'$ represents a stochastic component that can be either I(0) or I(1) process. Suppose a simple bivariate linear trend model given by

$$\begin{aligned} y_{1t} &= c_1 + \mu_1 t + \varepsilon_{1t}, \\ y_{2t} &= c_2 + \mu_2 t + \varepsilon_{2t}, \end{aligned}$$

where ε_{1t} and ε_{2t} are zero mean I(0) error terms, $\mu_1 \neq 0$ and $\mu_2 \neq 0$. Then, this model has a representation (1) with

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= \varepsilon_{1t}, \quad \text{and} \quad s_{2t} = \varepsilon_{2t}. \end{aligned} \tag{2}$$

According to the definition of Engle and Kozicki (1993), a feature is said to be common if a linear combination of the series fails to have the feature. Since the deterministic trend is the main feature of interest, two variables are cotrended if the trend is eliminated by taking a particular linear combination (see also Bierens, 2000, Hatanaka, 2000, and Hatanaka and Yamada, 2003). In the case of a linear deterministic trend in (2), there is a trivial cotrending relationship since the vector $(1, -\mu_1/\mu_2)$ can eliminate the trend. Likewise, if m variables are generated from a multivariate linear trend model, there are $m-1$ trivial cotrending relationships since there are $m-1$ linearly independent non-zero cotrending vectors.

In our analysis, stochastic trends can be either included or excluded. When stochastic trends are present, there will be two layers of potential cotrending relationships. For example, suppose a pair of variables are generated from two independent random-walk-with-drift processes:

$$\begin{aligned} y_{1t} &= \mu_1 + y_{1t-1} + \varepsilon_{1t}, \\ y_{2t} &= \mu_2 + y_{2t-1} + \varepsilon_{2t}, \end{aligned}$$

where ε_{1t} and ε_{2t} are zero mean iid error terms, $\mu_1 \neq 0$ and $\mu_2 \neq 0$. Then, the model has a representation (1) with

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= s_{1t-1} + \varepsilon_{1t}, \quad \text{and} \\ s_{2t} &= s_{2t-1} + \varepsilon_{2t} \end{aligned} \tag{3}$$

so that s_{1t} and s_{2t} are I(1) processes, or stochastic trends. In this case, the vector $(1, -\mu_1/\mu_2)$ eliminates the linear deterministic trend, but no linear combination can eliminate the stochastic trend. However, since the dominant trend, namely, the deterministic trend can still be eliminated, we refer to the vector $(1, -\mu_1/\mu_2)$ as a weak cotrending vector. In contrast, if (3) is replaced by

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= s_{1t-1} + \varepsilon_{1t}, \quad \text{and} \\ s_{2t} &= (\mu_2/\mu_1) s_{1t} + \varepsilon_{2t}, \end{aligned} \tag{4}$$

the weak cotrending vector $(1, -\mu_1/\mu_2)$ eliminates not only the linear deterministic trend, but also the stochastic trend. Since both type of trends are eliminated by a single vector $(1, -\mu_1/\mu_2)$, we view such a case as the stronger version of the cotrending relationship.

In a system of m variables with both stochastic and deterministic trends, one of our goals is to identify the total number of linearly independent vectors that can eliminate both stochastic and deterministic trends at the same time. In this paper, we refer to the number of such cotrending vectors as the cotrending rank and denote it by r_1. The cotrending rank can be any integer value in the range of $0 \leq r_1 < m$. In addition to r_1, we also introduce the weak cotrending rank (denoted by r_2) as the total number of linearly independent vectors that can eliminate the deterministic trend, regardless of whether such vectors can eliminate the stochastic trend at the same time. Since all the cotrending vectors are also weak cotrending vectors, r_2 should satisfy $r_1 \leq r_2 < m$. While it is not the stronger version of cotrending rank based on a broader notion of trends, the identification of r_2 is also important in the presence of both stochastic and deterministic trends, since $m - r_2$ in the m-variable-system corresponds to the total number of common deterministic trends. In the above example of $m = 2$, a vector $(1, -\mu_1/\mu_2)$ can eliminate the deterministic trend regardless of the values of μ_1 and μ_2. Thus, the weak cotrending rank r_2 of both models (3) and (4) is 1. However, the cotrending rank r_1 is 0 for model (3) and is 1 for model (4). In this paper, we propose a simple procedure to identify both r_1 and r_2 in a system of m variables, in the presence of both stochastic and deterministic trends.

As discussed above, the elimination of the deterministic trend is of primary interest in our cotrend analysis. This differs from traditional cointegration analysis where elimination of the stochastic trend is its main interest, even if a deterministic trend is included in the system. To see this point, consider another model with stochastic trends given by

$$\begin{aligned} d_{1t} &= c_1 + \mu_1 t, \quad d_{2t} = c_2 + \mu_2 t, \\ s_{1t} &= s_{1t-1} + \varepsilon_{1t}, \quad \text{and} \\ s_{2t} &= s_{1t} + \varepsilon_{2t}. \end{aligned} \tag{5}$$

Here the cointegrating vector $(1, -1)$ can always eliminate the stochastic trend, but not the deterministic trend unless $\mu_1 = \mu_2$. For the purpose of distinguishing between (4) and (5) in cointegration analysis, Ogaki and Park (1997) introduced the notions of stochastic cointegration and deterministic cointegration. In their terminology, stochastic cointegration refers to the case in which only the stochastic trend is eliminated by the cointegrating vector. In contrast, deterministic cointegration refers to the case in which both stochastic and deterministic trends are eliminated by the same cointegrating vector. In our cotrend analysis, however, two models differ because the (strong) cotrending rank r_1 is 1 for (4) but 0 for (5).

2.2 Trend breaks and smooth transition trends

So far, we have seen only an obvious cotrending relationship with a linear trend for the purpose of introducing the notion of cotrending ranks. However, cotrend analysis becomes more meaningful when variables contain various forms of nonlinear deterministic trends so that the system can have more than one common deterministic trend. Here, we provide some examples of nonlinear trends to highlight the class of deterministic trends that are allowed in our consistent cotrending rank selection procedure.

As discussed in Mills (2003), many macroeconomic time series data, including GDP of the UK and Japan and stock prices in the U.S., violate the assumption of stable growth over typical sample periods. A convenient approach to allowing for multiple shifts in the growth rate, while maintaining the continuity of the trend function, is to consider a kinked trend, or a piece-wise linear trend structure in each segment of the whole sample period. When there are h time shifts in the (log) growth rate, the segmented linear trend can be written as

$$d_t^{KINK} = \mu_0 t + \sum_{i=1}^{h} \mu_i (t - T_i) 1[t > T_i],$$

where T_i is the trend break point and $1[x]$ is an indicator that takes the value of 1 if x is true and 0, otherwise. The segmented linear trend implies that the growth rate corresponds to μ_0, during the first subperiod $t < T_1$, and corresponds to $\mu_0 + \sum_{i=1}^{j} \mu_i$, in the remaining subperiods, $T_j \leq t < T_{j+1}$ for $j = 1, ..., h$.

Recall that in the preceding bivariate example with a linear trend, the deterministic trend terms d_{1t} and d_{2t} are by definition proportional to a common linear deterministic trend, say $d_t^{LIN} = t$, ignoring the constant. Therefore, we can always find at least one linear combination that eliminates the trend, and the cotrending relationship is trivial. However, if the linear trend functions in d_{1t} and d_{2t} are replaced by segmented trend functions, a linear combination can eliminate the deterministic trend if and only if (i) all the break points, T_i's, are the same and (ii) all the piece-wise trend slope coefficients, μ_i's, are proportional between the two trend functions. If either of the two conditions fails to hold, the two nonlinear deterministic trends are linearly independent and no common deterministic trend exists. This fact also shows how our cotrend analysis differs from the cobreaking analysis of Hendry and Mizon (1998) and Clements and Hendry (1999). In the presence of a trend break, cobreaking is a necessary condition of cotrending, but not a sufficient condition.

Although the segmented trend function d_t^{KINK} imposes continuity, its first derivative is not continuous, suggesting an abrupt change of the growth rate at each break point. To allow for a gradual change in the growth rate, we may replace the indicator function in d_t^{KINK} with a smooth transition function. This substitution of the trend function leads to a smooth transition trend model. The smooth transition trend model was originally proposed by Bacon and Watts (1971) and has been discussed by Lin and

Teräsvirta (1994) and Leybourne, Newbold and Vougas (1998). While there are many types of smooth transition trend functions, one most frequently used is the logistic transition function given by

$$G(\gamma_i, T_i) = \frac{1}{1 + \exp(-\gamma_i(t - T_i))},$$

where γ_i (> 0) is the scaling parameter that controls the speed of transition, and T_i becomes the timing of the transition midpoint instead of the break point. The nonlinear deterministic trend component of a multiple-regime logistic smooth transition trend (LSTT) model takes the form of

$$d_t^{LST} = \mu_0 t + \sum_{i=1}^{h} \mu_i (t - T_i) G(\gamma_i, T_i).$$

It should be noted that, as γ_i approaches infinity, the logistic transition function $G(\gamma_i, T_i)$ approaches the indicator function $1[t > T_i]$. Thus, the deterministic trend d_t^{LST} nests both the kinked trend d_t^{KINK} and the linear trend d_t^{LIN} as special cases. Figure 1 shows the typical shape of kinked and smooth transition trends when $h = 1$. The former contains a one-time abrupt change in the first derivative, while the latter shows continuous change in the first derivative.

Both segmented and smooth transition type models of trend shift are allowed in our cotrending rank selection procedure. Furthermore, other types of nonlinear deterministic trend functions can be also included, as long as they belong to a class of trend functions so that their order of magnitude is identical to that of a linear trend. Let $\{d_t^{KINK}\}_{t=1}^{T}$, and $\{d_t^{LST}\}_{t=1}^{T}$ be the deterministic sequences where $T_i = k_i T$, $0 < k_0 < k_1 < ... < k_h < 1$ and γ_i's are fixed. Then, both trend sequences have the same order of magnitude as the linear trend sequence $\{d_t^{LIN}\}_{t=1}^{T}$ in the sense that both $\sum_{t=1}^{T} d_t^{KINK} / \sum_{t=1}^{T} d_t^{LIN}$ and $\sum_{t=1}^{T} d_t^{LST} / \sum_{t=1}^{T} d_t^{LIN}$ approach a non-zero constant as T tends to infinity. Similarly, our analysis remains valid for any nonlinear deterministic trend sequence $\{d_t^{*}\}_{t=1}^{T}$, such that $\sum_{t=1}^{T} d_t^{*} / \sum_{t=1}^{T} d_t^{LIN}$ approaches some non-zero constant as T tends to infinity. In the following section, we propose a procedure to identify both r_1 and r_2 in a system of m variables, which is valid for any nonlinear deterministic trend functions that belong to this class of nonlinear trends.[2] An important feature of our procedure is that estimation of parametric nonlinear trend functions is not required. In this sense, our procedure can be viewed as a nonparametric approach to cotrending rank selection.

3 Theory

We assume that an m-variate time series, $y_t = [y_{1t}, \cdots, y_{mt}]'$, is generated by

$$y_t = d_t + s_t, \quad t = 1, \cdots T, \tag{6}$$

where $d_t = [d_{1t}, \cdots, d_{mt}]'$ is a nonstochastic trend component, $s_t = [s_{1t}, \cdots, s_{mt}]'$ is a stochastic process, respectively defined below, and neither d_t nor s_t is observable. We denote a random (scalar) sequence x_T by $O_p(T^{\lambda})$ if $T^{-\lambda}x_T$ is bounded in probability, and by $o_p(T^{\lambda})$ if $T^{-\lambda}x_T$ converges to zero in probability. For a deterministic sequence, we use $O(T^{\lambda})$ and $o(T^{\lambda})$, if $T^{-\lambda}x_T$ is bounded and converges to zero, respectively. The first difference of x_t is denoted by Δx_t. Below, we employ a set of assumptions that are similar to those in Hatanaka and Yamada (2003).

Assumption 1. (i) $s_t = s_{t-1} + \xi_t$ and $\xi_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}$, $C_0 = I_n$, $\sum_{j=0}^{\infty} j^2 \|C_j\| < \infty$, where ε_t is iid with zero mean and covariance matrix $\Sigma_{\varepsilon\varepsilon} > 0$. (ii) Each element of $\sum_{t=1}^{T} d_t$ is $O(T^2)$ and is not $o(T^2)$. (iii) There exists an $m \times m$ orthogonal full rank matrix $B = [B_{\perp} \ B_2 \ B_1 \ ]$, such that each element of $\sum_{t=1}^{T} B_1' y_t$ is $O_p(T^{1/2})$, each element of $\sum_{t=1}^{T} B_2' y_t$ is $O_p(T)$ and is not $o_p(T)$, and each element of $\sum_{t=1}^{T} B_{\perp}' y_t$ is $O_p(T^2)$ is not $o_p(T^2)$, where B_1, B_2, $B_{\perp}$ are $m \times r_1$, $m \times (r_2 - r_1)$ and $m \times (m - r_2)$, respectively.

[2] We focus on this class of trends since the trend breaks are the most frequently used forms of nonlinear trends in practice. However, we can easily extend our approach to incorportate other classes of trend functions, such as the one for quadratic trends or cubic trends.

Under Assumption 1, B_1 represents a set of cotrending vectors that eliminates both deterministic and stochastic trends. B_2 represents a set of vectors eliminating only deterministic trends, but not stochastic trends. $B_\perp$ consists of vectors orthogonal to B_1 and B_2.

In the scalar case, the von Neumann ratio is defined as the ratio of the sample second moment of the differences to that of the level of a time series. The multivariate generalization of the von Neumann ratio is defined as $S_{11}^{-1}S_{00}$ where

$$S_{11} = T^{-1}\sum_{t=1}^{T} y_t y_t^{'}, \quad \text{and} \quad S_{00} = T^{-1}\sum_{t=2}^{T} \Delta y_t \Delta y_t^{'}.$$

Shintani (2001) and Harris and Poskitt (2004) also use this multivariate version of the von Neumann ratio in cointegration analysis. Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_m \geq 0$ be the eigenvalues of $S_{11}^{-1}S_{00}$. We summarize the statistical properties of $\widehat{\lambda}_i^{'}$s in the presence of both stochastic and deterministic trends in the following lemma.

Lemma 1 *Under Assumption 1, we have: (i) a sequence of* $[\widehat{\lambda}_1, \cdots, \widehat{\lambda}_{r_1}]$ *has a positive limit and is* $O_p(1)$ *but is not* $o_p(1)$*; (ii) a sequence of* $T[\widehat{\lambda}_{r_1+1}, \cdots, \widehat{\lambda}_{r_2}]$ *has a positive limit and is* $O_p(1)$ *but is not* $o_p(1)$*, provided* $r_2 - r_1 > 0$*; and (iii) a sequence of* $T^2[\widehat{\lambda}_{r_2+1}, \cdots, \widehat{\lambda}_m]$ *has a positive limit and is* $O_p(1)$ *but is not* $o_p(1)$*, provided* $m - r_2 > 0$*.*

From Lemma 1, the eigenvalues of $S_{11}^{-1}S_{00}$ can be classified into three groups depending on their rates of convergence, namely, $O_p(1)$, $O_p(T^{-1})$ and $O_p(T^{-2})$. The number of eigenvalues in each group corresponds to the number of cotrending relationships (r_1), the difference between weak cotrending and (strong) cotrending relationships ($r_2 - r_1$) and the number of common deterministic trends ($m - r_2$), respectively.[3] We exploit this property to construct the following two types of consistent cotrending

[3] The von Neumann ratio can be computed using demeaned series instead of using raw series. Unlike cotrending rank tests that require the limiting distribution of the von Neumann ratio, our procedure relies only on the convergence rate of its eigenvalues. Therefore, all of our theoretical results hold for the demeaned version of the von Neumann ratio.

rank selection procedures based on the von Neumann criterion, which is defined as a sum of the partial sum of eigenvalues and a penalty term. The first is a 'paired' procedure that independently selects the cotrending rank r_1 and the weak cotrending rank r_2 by minimizing each of

$$\begin{aligned} VN_1(r_1) &= -\sum_{i=1}^{r_1} \widehat{\lambda}_i + f(r_1)\frac{C_T}{T}, \text{ and} \\ VN_2(r_2) &= -\sum_{i=1}^{r_2} \widehat{\lambda}_i + f(r_2)\frac{C_T'}{T^2}, \end{aligned}$$

or

$$\begin{aligned} \widehat{r}_1 &= \arg \min_{0 \le r_1 \le m} VN_1(r_1), \text{ and} \\ \widehat{r}_2 &= \arg \min_{0 \le r_2 \le m} VN_2(r_2) \end{aligned}$$

where $f(r)$, C_T and C_T' are elements of the penalty function defined in detail below.

The second procedure is a 'joint' procedure that simultaneously determines both r_1 and r_2 by minimizing

$$VN(r_1, r_2) = -\sqrt{T}\sum_{i=1}^{r_1} \widehat{\lambda}_i - \sum_{i=r_1+1}^{r_2} \widehat{\lambda}_i + f(r_1)\frac{C_T}{T} + f(r_2)\frac{C_T'}{T^2},$$

or

$$(\widehat{r}_1, \widehat{r}_2) = \arg \min_{0 \le r_1, r_2 \le m} VN(r_1, r_2).$$

The main theoretical result is provided in the following proposition.

Proposition 1 *(i) Suppose Assumption 1 holds, and $f(r)$ is an increasing function of r, $C_T, C_T' \to \infty$, $C_T/T, C_T'/T, \to 0$, then the paired procedure using $VN_1(r_1)$ and $VN_2(r_2)$ yields,*

$$\lim_{T \to \infty} P(\widehat{r}_1 = r_1, \widehat{r}_2 = r_2) = 1.$$

(ii) Suppose Assumption 1 holds, and $f(r)$ is an increasing function of r, $C_T/\sqrt{T}, C_T'/\sqrt{T} \to \infty$, $C_T/T, C_T'/T, \to 0$, then the joint procedure using $VN(r_1, r_2)$ yields,

$$\lim_{T\to\infty} P(\widehat{r}_1 = r_1, \widehat{r}_2 = r_2) = 1.$$

Remarks:

(a) The proposition shows that both of the two cotrending rank selection procedures are consistent in selecting a cotrending rank without specifying a parametric model as long as the trend belongs to a certain class of nonlinear functions. The joint selection procedure requires slightly stronger assumptions on C_T and $C_T^{'}$ than the paired selection procedure.

(b) Commonly employed C_T in the literature of information criteria includes $C_T = \ln(T)$, $2\ln(\ln(T))$, and 2, which respectively lead to the Bayesian information criterion (BIC), Hannan-Quinn criterion (HQ), and Akaike information criterion (AIC). Part (i) of the proposition implies that the paired cotrending rank selection procedure is consistent when BIC and HQ type penalties are employed, but is inconsistent when an AIC type penalty is employed. In contrast, part (ii) of the proposition implies that C_T (and $C_T^{'}$) should diverge at a rate faster than $\sqrt{T}$ for the joint cotrending rank selection procedure, thus none of $C_T = \ln(T)$, $2\ln(\ln(T))$, and 2 yield consistency.

(c) By the definition of $VN(r_1, r_2)$, cotrending ranks selected by the joint procedure always satisfy $\widehat{r}_1 \leq \widehat{r}_2$. For the paired procedure, selected cotrending ranks will satisfy $\widehat{r}_1 \leq \widehat{r}_2$ if $C_T' = T^{\alpha} C_T$, where $0 \leq \alpha < 1$. This fact can be demonstrated by the following argument. The selected cotrending rank $\widehat{r}_1$ implies that $VN_1(r) > VN_1(\widehat{r}_1)$ for all $r < \widehat{r}_1$. This result is equivalent to the partial sum of eigenvalues $\sum_{i=r+1}^{\widehat{r}_1} \widehat{\lambda}_i$ being greater than $\{f(\widehat{r}_1) - f(r)\} C_T T^{-1}$ (note that $\widehat{\lambda}_i \geq 0$ and $f(\widehat{r}_1) - f(r) > 0$). To see if $VN_2(r) > VN_2(\widehat{r}_1)$ for the corresponding r and $\widehat{r}_1$, it suffices to show that $\sum_{i=r+1}^{\widehat{r}_1} \widehat{\lambda}_i$ is greater than $\{f(\widehat{r}_1) - f(r)\} C_T' T^{-2}$. By substituting $C_T' = T^{\alpha} C_T$, the latter becomes $\{f(\widehat{r}_1) - f(r)\} C_T T^{-1} \times$

$T^{-(1-\alpha)}$. Since $T^{-(1-\alpha)} < 1$, $\sum_{i=r+1}^{\widehat{r}_1} \widehat{\lambda}_i > \{f(\widehat{r}_1) - f(r)\} C_T T^{-1} > \{f(\widehat{r}_1) - f(r)\} C_T T^{-1} \times T^{-(1-\alpha)}$. Because we have shown that $VN_2(r) > VN_2(\widehat{r}_1)$ for all $r < \widehat{r}_1$, it implies $\widehat{r}_1 \leq \widehat{r}_2$.

(d) The criterion function $VN_1(r_1)$ in the paired procedure can solely be used to select cointegrating rank in a system of stochastic trends without nonlinear deterministic trends. It nests the criterion function considered in Harris and Poskitt (2004) as a special case. Their criterion $\Gamma_{C,T}$, in their notation, is identical to $VN_1(r_1)$ combined with $C_T = \ln(T)$ and $f(r) = 2r(2m - r + 1)$. Thus, part (i) of the proposition extends the result of Harris and Poskitt (2004) to the cointegrating rank selection for general choice of C_T and $f(r)$.

(e) For consistency of our procedures, $f(r)$ can be any increasing function of r. In this paper, we follow Harris and Poskitt (2004) and employ $f(r) = 2r(2m - r + 1)$, the function used in their consistent cointegrating rank selection criterion. This choice satisfies the required condition of an increasing function since $df(r)/dr = 4(m - r) > 0$. Other choices of function, such as $f(r) = 2mr - r^2$ and $f(r) = 2mr - r(r + 1)/2$, are also discussed in Cheng and Phillips (2009) based on the reduced rank regression structure of the cointegrated system.

4 Experimental evidence

4.1 Stochastic trends and cointegrating rank

The proposed cotrending rank selection procedures are justified according to the asymptotic theory. Thus, it is of interest to examine their finite sample properties by means of Monte Carlo analysis. This section reports the results under different settings of the true cotrending ranks, and of various penalty terms.

Before we present the main simulation results of cotrending rank selection in a system with stochastic and nonlinear deterministic trends, let us first consider the case of a cointegrated system without deterministic trends. Understanding the basic characteristics of the multivariate von Neumann ratio-based procedure in a simple system with stochastic trends only, will help us justify the use of the similar procedure in a more complicated system. Recall that the von Neumann ratio criterion $VN_1(r_1)$ in the paired procedure can be used to determine the cointegrating rank in the cointegrated system, and that it nests the cointegrating rank selection procedure of Harris and Poskitt (2004) as a special case. Since estimation of the cointegrating vector and serial correlation structure is not required, our procedure and the procedure by Harris and Poskitt (2004) may be viewed as a nonparametric approach to cointegrating rank selection. In contrast, the information criteria for selecting cointegrating rank in Cheng and Phillips (2009) are based on the eigenstructure of a reduced rank regression model. While serial correlation structure is not estimated, cointegrating vectors are estimated. In this sense, their procedure may be viewed as a semiparametric approach to cointegrating rank selection. Here, we use the same simulation design as in Cheng and Phillips (2009) and compare the finite sample performance of two alternative approaches.

A bivariate time series $y_t = (y_{1t}, y_{2t})'$ is generated from

$$\Delta y_t = \alpha\beta' y_{t-1} + u_t, , \quad t = 1, \cdots T,$$

where u_t follows a $VAR(1)$ process with a VAR coefficient $0.4 \times I_2$ and a mutually independent standard normal error term. By setting $\alpha\beta' = 0$,

$$\alpha\beta' = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix},$$

and

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

we generate a multivariate system with the true cointegrating rank $r_1 = 0, 1$ and 2, respectively. We evaluate the finite sample performance of both semiparametric and nonparametric approaches by the frequencies of selecting the true cointegrating rank in 20,000 replications for the sample sizes $T = 50$, 100 and 400.[4] For the reduced rank regression procedure of Cheng and Phillips (2009), we employ the AIC, BIC and HQ criteria and denote them by RRR-AIC, RRR-BIC and RRR-HQ, respectively. The von Neumann ratio criterion $\Gamma_{C,T}$ of Harris and Poskitt (2004) is equivalent to $VN_1(r_1)$ with $f(r) = 2r(2m - r + 1)$ and $C_T = \ln(T)$. Since it involves a BIC-type penalty, we refer to this procedure by VN-BIC. In addition, we also consider the AIC-type penalty $C_T = 2$, as well as an HQ type penalty $C_T = 2\ln(\ln(T))$, and denote corresponding criteria by VN-AIC and VN-HQ, respectively. It should be noted that theoretical analysis implies that both RRR-AIC and VN-AIC are inconsistent in selecting true cointegrating rank.

Table 1 reports the performance of the cointegrating rank selection procedures based on six criteria, with frequencies of correctly selecting true rank shown in bold fonts. The results of the simulation can be summarized as follows.

First, the semiparametric approach by Cheng and Phillips (2009) and our nonparametric approach seem to complement each other because their relative performance depends on the data generating processes. If true cointegrating rank is $r_1 = 0$, the nonparametric von Neumann ratio-based procedures uniformly outperform the semiparametric reduced rank regression-based procedures for all the sample sizes under consideration. In contrast, if the true cointegrating rank is $r_1 = 2$ and the sample size is small ($T = 50$ and 100), each of the reduced rank regression procedures, RRR-AIC, RRR-BIC and RRR-HQ, works better than each counterpart of the von Neumann ratio procedures, VN-AIC, VN-BIC and VN-HQ, respectively. If the true cointegrating rank is $r_1 = 1$, the semiparametric reduced rank regression procedure works better with a BIC type penalty (RRR-BIC) when the sample size is as small as $T = 50$, but the nonparametric von Neumann ratio procedures dominates for the other cases.

[4] Here, we follow Cheng and Phillips (2009) and the first 50 observations are discarded to eliminate the effect of the initial values $y_0 = (0, 0)'$ and $u_0 = (0, 0)'$.

Second, for the von Neumann ratio-based procedures, the AIC type penalty often works well when the sample size is small, despite the fact that it provides theoretically inconsistent rank selection. In particular, it dominates other types of penalties if the true cointegrating rank is the largest ($r_1 = 2$), mainly because the penalty for higher rank is much smaller with $C_T = 2$ than with $C_T = \ln(T)$ or $C_T = 2\ln(\ln(T))$. However, even in the case of low frequencies of selecting the true rank when the sample size is small, they quickly approach one when the sample size increases to $T = 400$. On the whole, it seems fair to say that the von Neumann criterion is at least as useful as the information criterion based on the reduced rank regression in selecting cointegrating rank.

4.2 Deterministic trends and cotrending rank

In this subsection, we evaluate the finite sample performance of our proposed procedure using the three-dimensional vector series $y_t^* = (y_{1t}^*, y_{2t}^*, y_{3t}^*)'$ with different combinations of cotrending and weak cotrending ranks ($m = 3$).

To consider the case with only one common (nonlinear) deterministic trend, we first generate the data using

$$\begin{aligned} y_{1t}^* &= \rho_1 y_{1t-1}^* + \varepsilon_{1t}, \\ y_{2t}^* &= \rho_2 y_{2t-1}^* + \varepsilon_{2t}, \\ y_{3t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau T \\ c + (\mu_0 - \mu_1)\tau T + \mu_1 t & \text{if } t > \tau T \end{cases}, \end{aligned} \tag{7}$$

with $(\varepsilon_{1t}, \varepsilon_{2t})' = iidN(0, \Sigma_\varepsilon)$ where

$$\Sigma_\varepsilon = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Note that here y_{1t}^* and y_{2t}^* do not have a deterministic trend, but the transformed system becomes equivalent to y_t given in equation (6), where each element contains a deterministic trend and a stochastic component. We can use any nonsingular matrix A such that $y_t = Ay_t^* = d_t + s_t$. Because the eigenvalues for the von Neumann ratio are invariant to any nonsingular transformation of the data, we can directly use y_t^* in the computation of our rank selection criteria in place of $y_t = Ay_t^*$ in the simulation. For example, a transformation using a matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

yields

$$\begin{aligned} y_{1t} &= y_{1t}^* + y_{2t}^* + y_{3t}^* = d_{1t} + s_{1t} \\ y_{2t} &= -y_{1t}^* + y_{2t}^* + y_{3t}^* = d_{2t} + s_{2t} \\ y_{3t} &= y_{1t}^* + y_{3t}^* = d_{3t} + s_{3t}. \end{aligned}$$

In the case of $\rho_1 = 0.5$ and $\rho_2 = 1.0$, a vector $(1, -1, 0)$ becomes a cotrending vector since $y_{1t} - y_{2t} = 2y_{1t}^*$ is stationary, and a vector $(1, 0, -1)$ becomes a weak cotrending vector since $y_{1t} - y_{3t} = y_{2t}^*$ contains a stochastic trend but not a deterministic trend. Since other cotrending vectors can be also incorporated by a different choice of a nonsingular matrix A, a very large class of cotrended systems can be covered by our simple simulation design.

We consider three cases by using different combinations of $\rho_i \in \{0.5, 1.0\}$ for $i = 1, 2$, in (7) and generate the data with $(r_1, r_2) = (2, 2)$, $(1, 2)$ and $(0, 2)$. In particular, setting $\rho_1 = \rho_2 = 0.5$ implies $(r_1, r_2) = (2, 2)$, $\rho_1 = 0.5$ and $\rho_2 = 1.0$ implies $(r_1, r_2) = (1, 2)$ and $\rho_1 = \rho_2 = 1.0$ implies $(r_1, r_2) = (0, 2)$. The parameters for the kinked trend function are set to $c = 0.5$, $\mu_0 = 2$, $\tau = 0.5$ and $\mu_1 = 0.5$.

Second, we consider the cases of two deterministic trends using

$$\begin{aligned} y_{1t}^* &= \rho_1 y_{1t-1}^* + \varepsilon_{1t}, \\ y_{2t}^* &= c + \mu_0 t \\ y_{3t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau T \\ c + (\mu_0 - \mu_1)\tau T + \mu_1 t & \text{if } t > \tau T \end{cases}, \end{aligned} \tag{8}$$

with $\varepsilon_{1t} = iidN(0,1)$, $\rho_1 \in \{0.5, 1.0\}, c = 0.5$, $\mu_0 = 2$ and $\tau = 0.5$. This system generates the data with $(r_1, r_2) = (1, 1)$ when $\rho_1 = 0.5$, and $(r_1, r_2) = (0, 1)$ when $\rho_1 = 1.0$.

Finally, we consider the three-deterministic trend case using

$$\begin{aligned} y_{1t}^* &= c + \mu_0 t + \varepsilon_{1t}, \\ y_{2t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau_1 T \\ c + (\mu_0 - \mu_1)\tau_1 T + \mu_1 t & \text{if } t > \tau_1 T \end{cases} \\ y_{3t}^* &= \begin{cases} c + \mu_0 t & \text{if } t \leq \tau_2 T \\ c + (\mu_0 - \mu_1)\tau_2 T + \mu_1 t & \text{if } t > \tau_2 T \end{cases}, \end{aligned} \tag{9}$$

with $\varepsilon_{1t} = iidN(0,1)$, $c = 0.5$, $\mu_0 = 2$, $\tau_1 = 0.5, \tau_2 = 1/3$ and $\mu_1 = 0.5$. This system generates the data with $(r_1, r_2) = (0, 0)$.

We employ two paired cotrending rank selection procedures and two joint cotrending rank selection procedures. For the paired procedures, we use a BIC type penalty $C_T = \ln(T)$ for $VN_1(r_1)$. Recall that selected cotrending ranks from the paired procedure always satisfy $\widehat{r}_1 \leq \widehat{r}_2$ as long as $C_T' = T^\alpha C_T$, where $0 \leq \alpha < 1$. Here, we employ $C_T' = \sqrt{T}\ln(T)$ for $VN_2(r_2)$ and denote corresponding paired procedure by 'paired BIC.' In addition, we also consider the case with a weaker penalty for $VN_2(r_2)$ by replacing the penalty with $C_T' = \sqrt{T}\ln(\ln(T))$. Since $VN_1(r_1)$ is the same as before but the penalty for $VN_2(r_2)$ somewhat resembles that of the HQ type penalty, we denote the procedure by 'paired BIC-HQ.'

For the joint selection procedures, $\ln(T)$ cannot be used for C_T(and C_T'), since consistency requires the penalty to diverge at a rate faster than $\sqrt{T}$. Therefore, we consider $VN(r_1, r_2)$ with the penalty $C_T = C'_T = \sqrt{T}\ln(T)$, and denote the procedure as 'joint BIC.' We additionally consider the pair of slower rate $C_T = C'_T = \sqrt{T}\ln(\ln(T))$ and denote the corresponding procedure as 'joint HQ.' As in the case of cointegration analysis, we employ $f(r) = 2r(2m - r + 1)$.

Tables 2 to 4 report the frequencies of selecting cotrending rank r_1 and weak cotrending rank r_2 by four procedures for sample sizes $T = 50$, 100 and 400 in 20,000 replications.[5] For each data generating process, the pair $(\widehat{r}_1, \widehat{r}_2)$ is selected by minimizing the von Neumann criterion among $(r_1, r_2) = (2, 2)$, $(1, 2)$, $(0, 2)$, $(1, 1)$, $(0, 1)$ and $(0, 0)$.[6] Frequencies of selecting the true model are shown in a bold font in the table. The results of the simulation can be summarized as follows.

First, both the paired procedures and joint procedures work well even when the sample size is as small as $T = 50$. When there is only one common deterministic trend and $T = 50$, paired procedures, paired BIC and paired BIC-HQ, work better than the joint procedures, joint BIC and joint HQ, for the cases $(r_1, r_2) = (2, 2)$ and $(1, 2)$, but the latter works better for the case of $(r_1, r_2) = (0, 2)$. However, as the sample size increases, the frequencies of selecting the true rank become close to one for both types of procedures, and thus the performance of the two procedures becomes almost indistinguishable.

Second, when there are two common deterministic trends and $T = 50$, the paired procedures perform better for the case of $(r_1, r_2) = (1, 1)$ and the joint procedures perform better for the case of $(r_1, r_2) = (0, 1)$. When $T = 100$, both procedures yield sufficiently high frequencies of selecting the true rank.

Finally, when there are three deterministic trends in the system, or $(r_1, r_2) = (0, 0)$, the performance highly depends on the choice of penalty terms. In particular, paired BIC and joint BIC select true rank all the time even if the sample size is $T = 50$. In contrast, the frequencies are very low for the paired BIC-HQ and joint HQ when the sample size is small ($T = 50$), and frequencies become close to unity

[5] For the stationary AR(1) part of the equations, the initial values are generated from its stationary distribution. For the other equations, initial values are set at 0.

[6] We only report the results from raw series version of the von Neumann criterion in the simulation since the demeaned version yielded similar results. The full simulation results are available upon request.

only when the sample size is $T = 400$.

4.3 Smooth transition trends and cotrending rank

In this section, we study the effect of nonlinearity in the trend function on the performance of our cotrending rank selection procedure. To this end, we consider the logistic smooth transition trend model, and control the shape of the deterministic function by controlling the scale parameters in the logistic transition function. We generate the artificial data with $(r_1, r_2) = (0, 1)$ using

$$\begin{aligned} y^*_{1t} &= y^*_{1t-1} + \varepsilon_{1t}; \\ y^*_{2t} &= c_0 + \mu_0 t; \\ y^*_{3t} &= (c_0 + \mu_0 t)G(\gamma, \tau T) + (c_1 + \mu_1 t)(1 - G(\gamma, \tau T)) \end{aligned} \tag{10}$$

where $G(\gamma, \tau T)$ is a logistic transition function defined in section 2 and $\varepsilon_{1t} = iidN(0,1)$, $c = 0.5$, $\mu_0 = 2$, $\tau = 0.5$ and $\mu_1 = 0.5$. As noted in section 2, the scale parameter γ controls the speed of transition. As γ approaches infinity, the logistic function collapses to an index function $I(t > \tau T)$ and (10) become (8) with $\rho_1 = 1.0$. On the other hand, as γ approaches zero, the smooth transition trend model approaches to a linear trend. In this scenario, we can always find the linear combination that eliminates the trend function. In other words, when γ is close to zero, the system of two common deterministic trends ($r_2 = 1$) becomes closer to the system of one common deterministic trend ($r_2 = 2$). Therefore, for a small value of γ, we expect that it will be difficult for our procedure to identify $r_2 = 1$ from $r_2 = 2$.

Table 5 presents the simulation results given different choices of the scale parameter $\gamma \in \{0.001, 0.005, 0.01\}$ when $T = 400$. Note that we can use the result of Table 4 for $(r_1, r_2) = (0, 1)$ and $T = 400$ as the benchmark limit case with a large γ. It turns out that the procedure works well in selecting the true rank even γ is as small as 0.01. Consistent with our prediction, two of the four procedures (joint BIC-HQ and joint HQ) select $(r_1, r_2) = (0, 2)$ when $\gamma = 0.005$, and other procedures select $(r_1, r_2) = (0, 2)$ with high frequencies when $\gamma = 0.001$.

5 Application

The simulation results in the previous section show that our procedures perform well in various experimental set-ups. In this section, we apply our procedures to the Japanese money demand function to investigate the cotrending relations among money demand, income and interest rate ($m = 3$). A seasonally adjusted quarterly series of real GDP, two definitions of monetary aggregates, $M1$ and $M2$, and the call rate for the sample period from 1980:Q1 to 2010:Q4, are plotted in Figures 2 to 5. The figures show the possibility of kinked deterministic trends in these variables.

We follow Bae, Kakkar and Ogaki (2006) and consider following three different specifications of money demand functions,

$$\begin{aligned}
\text{Model 1} \quad &: \quad \ln\left(\frac{M_t}{P_t}\right) = \beta_0 + \beta_1 \ln(y_t) + \beta_1 i_t + \varepsilon_t, \\
\text{Model 2} \quad &: \quad \ln\left(\frac{M_t}{P_t}\right) = \beta_0 + \beta_1 \ln(y_t) + \beta_1 \ln(i_t) + \varepsilon_t, \text{ and} \\
\text{Model 3} \quad &: \quad \ln\left(\frac{M_t}{P_t}\right) = \beta_0 + \beta_1 \ln(y_t) + \beta_1 \ln\left(\frac{i_t}{1+i_t}\right) + \varepsilon_t,
\end{aligned}$$

where M_t is the money demand, P_t is the aggregate price level, y_t is real GDP and i_t is the nominal interest rate.

We apply both paired and joint cotrending rank selection procedures to the vectors $(\ln(M_t/P_t), \ln(y_t), i_t)$, $(\ln(M_t/P_t), \ln(y_t), \ln(i_t))$, and $(\ln(M_t/P_t), \ln(y_t), \ln(i_t/(1+i_t)))$. Table 6 reports the empirical results for all three different specifications of the functional form for interest elasticity of money demand. The results are somewhat mixed depending on the choice of the penalty of the criteria and the choice of the variables. However, it is important to note that none of the procedures select $(r_1, r_2) = (0, 0)$. This implies that there are, at least, either cotrending or weak cotrending relationships in Japanese money demand in the long-run. When $M2$ is used as the monetary aggregate and when demeand version of the von Neumann ratio is used, $(r_1, r_2) = (0, 2)$ is selected for all cases, implying that the kinked trend is likely to be a single common deterministic trend among three variables.

6 Conclusion

This paper has proposed a model-free cotrending rank selection procedure to use when both stochastic and nonlinear deterministic trends are present in a multivariate system. The procedure selects two types of cotrending ranks by minimizing two new criteria based on the generalized von Neumann ratio. Our approach is invariant to the linear transformation of data, robust to misspecification of the model and consistent under very general conditions. Monte Carlo experiments have suggested good finite sample performance of the proposed procedure. An empirical application to the money demand function in Japan has also suggested the usefulness of our procedure in detecting cotrending relationships when nonlinear deterministic trends are present in data.

Appendix

Proof of Lemma 1:

We want to show that $\widehat{\lambda}_1,\cdots,\widehat{\lambda}_{r_1}$ is $O_p(1)$ but is not $o_p(1)$, $\widehat{\lambda}_{r_1+1},\cdots,\widehat{\lambda}_{r_2}$ is $O_p(T^{-1})$ but is not $o_p(T^{-1})$, and $\widehat{\lambda}_{r_2+1}, \cdots, \widehat{\lambda}_m$ is $O_p(T^{-2})$ but is not $o_p(T^{-2})$ if all the eigenvalues of $S_{11}^{-1}S_{00}$ are arranged in a descending order. We employ the data matrix notation, $Y^{'} = [y_1, \cdots, y_T]$, $D^{'} = [d_1, \cdots, d_T]$ and $S^{'} = [s_1, \cdots, s_T]$.

We have constructed an orthogonal full rank matrix $[B_\perp \ B_2 \ B_1]$ in Assumption 1 and further define

$$M_{11} = B^{'}S_{11}B, \text{ and } M_{00} = B^{'}S_{00}B.$$

Due to the orthogonality of the matrix $[B_\perp \ B_2 \ B_1]$, the eigenvalues of $S_{11}^{-1}S_{00}$ arise as the same solutions to

$$\det(\lambda M_{11} - M_{00}) = 0.$$

Our proof can be established in the following two steps.

Step 1:

We assume $G = \lim_{T\to\infty} T^{-3}\sum_{t=1}^{T} d_t d_t'$ exists and $T^{-3}\sum_{t=1}^{T} d_t d_t' - G$ is $O(T^{-1/2})$. The eigenvalues of $T^2 M_{11}^{-1} M_{00}$ are equivalent to the eigenvalues $\lambda^{'}s$ that solve

$$\det(\lambda T^{-2} M_{11} - M_{00}) = 0.$$

For the matrix $T^{-2}M_{11}$, the only block matrix that is not equal to zero is $B_\perp^{'} Y^{'} Y B_\perp$, which converges to $B_\perp^{'} G B_\perp$ under Assumption 1. Because the eigenvalues are continuous functions of the matrix,

$$p\lim_{T\longrightarrow\infty} \lambda_i(T^2 M_{11}^{-1} M_{00}) = \lambda_i(p\lim_{T\longrightarrow\infty} T^2 M_{11}^{-1} M_{00}).$$

It can be easily shown that M_{00} is $O_p(1)$ but is not $o_p(1)$. Therefore, for $i = r_2 + 1, \cdots, m$, we are led to

$$\lambda_i(T^2 M_{11}^{-1} M_{00}) = O_p(1) \text{ but is not } o_p(1).$$

This outcome leads to the result that $T^2\widehat{\lambda}_i$ is $O_p(1)$ but not $o_p(1)$ for $i = r_2 + 1, \cdots, m$.

Step 2:

Let $D_T = diag[I_{m-r_2}, T^{1/2} I_{r_2}]$, the roots of

$$\det(\lambda T^{-2} M_{11} - M_{00}) = 0$$

are equivalent to

$$\det(D_T \left[\lambda T^{-2} M_{11} - M_{00}\right] D_T) = 0. \tag{11}$$

The matrix $\lambda T^{-2} M_{11}$ can be rewritten as

$$\begin{pmatrix} \lambda T^{-3} B_{\perp}^{'} Y^{'} Y B_{\perp} & \lambda T^{-3} B_{\perp}^{'} Y^{'} Y [\ B_2 \quad B_1\ ] \\ \lambda T^{-3} [\ \begin{matrix} B_2^{'} \\ B_1^{'} \end{matrix}\ ] Y^{'} Y B_{\perp} & \lambda T^{-3} [\ \begin{matrix} B_2^{'} \\ B_1^{'} \end{matrix}\ ] Y^{'} Y [\ B_2 \quad B_1\ ] \end{pmatrix},$$

and we denote

$$Y_a = \lambda T^{-3} B_{\perp}^{'} Y^{'} Y B_{\perp} - B_{\perp}^{'} \Delta Y^{'} \Delta Y B_{\perp}^{\prime},$$

$$Y_b = \lambda T^{-2} \begin{pmatrix} B_2^{'} Y^{'} Y B_2 & B_2^{'} Y^{'} Y B_1 \\ B_1^{'} Y^{'} Y B_2 & B_1^{'} Y^{'} Y B_1 \end{pmatrix} - \begin{pmatrix} T B_2^{'} \Delta Y^{'} \Delta Y B_2 & T^{1/2} B_2^{'} \Delta Y^{'} \Delta Y B_1 \\ T^{1/2} B_1^{'} \Delta Y^{'} \Delta Y B_2 & T B_1^{'} \Delta Y^{'} \Delta Y B_1 \end{pmatrix},$$

and

$$Y_c = \lambda T^{-\frac{5}{2}} \begin{pmatrix} B_2^{'} Y^{'} Y B_{\perp} \\ B_1^{'} Y^{'} Y B_{\perp} \end{pmatrix} - T^{1/2} \begin{pmatrix} B_2^{'} \Delta Y^{'} \Delta Y B_{\perp} \\ B_1^{'} \Delta Y^{'} \Delta Y B_{\perp} \end{pmatrix}.$$

Then equation (11) is rewritten as

$$\det(Y_a) \det[Y_b - Y_c^{'} Y_a^{-1} Y_c] = 0. \tag{12}$$

The first determinant can on the LHS of (12) cannot be equal to zero, implying the second determinant must be zero. Concerning the first part of Y_b,only its first $r_2 \times r_2$ diagonal block is nonzero, and the second part of Y_b and $Y_c^{'} Y_a^{-1} Y_c$ is $O_p(T)$ but is not $o_p(T)$. Hence, we are led to

$$\det(\lambda_i T^{-2} B_1^{'} Y^{'} Y B_1 - O_p(T)) = 0$$

for $i = r_1 + 1, \cdots, r_2$. While we let T goes to infinity and the solutions λ_i solves the above equation satisfies

$$\lambda_i(T^2 M_{11}^{-1} M_{00}) = O_p(T) \text{ but is not } o_p(T) \text{ for } i = r_1 + 1, \cdots, r_2.$$

Therefore, one can conclude that $\widehat{\lambda}_i$ is $O_p(T^{-1})$ but is not $o_p(T^{-1})$ for $i = r_1 + 1, \cdots, r_2$. Analogously, one can show that $\widehat{\lambda}_i$ is $O_p(1)$ but is not $o_p(1)$ for $i = 1, \cdots, r_1$.

Proof of Proposition 2.

(i) Let r_1 be the true cotrending rank, which is estimated by minimization of $VN_1(r_1)$ for $0 \leq r_1 \leq m$. To check the consistency of this estimator, we need to show $VN(r_1^{'}) > VN(r_1)$ if $r_1^{'}$ is not equal to the true cotrending rank r_1.

When $r_1^{'} < r_1$,

$$VN_1(r_1^{'}) - VN_1(r_1) = \sum_{i=r_1^{'}+1}^{r_1} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T T^{-1}.$$

In order to select r_1 with probability approaching 1 as $T \to \infty$, we need

$$\sum_{i=r_1^{'}+1}^{r_1} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T T^{-1} > 0, \text{ as } T \to \infty.$$

From Proposition 1, we know the first term is a positive number that is bounded away from zero and the second term is a negative number of order $O(C_T T^{-1})$. As long as $C_T T^{-1} \to 0$ as $T \to \infty$, the above inequality holds and we are led to the conclusion that $VN_1(r_1^{'}) > VN_1(r_1)$ when $r_1^{'} < r_1$.

When $r_1^{'} > r_1$,

$$VN_1(r_1^{'}) - VN_1(r_1) = -\sum_{i=r_1+1}^{r_1^{'}} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T T^{-1}.$$

From Proposition 1, we know that $\widehat{\lambda}_i$ is $O_p(T^{-1})$ but is not $o_p(T^{-1})$ for $i = r_1 + 1, \cdots r_2$, By multiplying both sides by T, we have

$$T\left(VN_1(r_1^{'}) - VN_1(r_1)\right) = -T\sum_{i=r_1+1}^{r_1^{'}} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))C_T.$$

As long as $C_T \to \infty$ as $T \to \infty$, the second term on the right hand side dominates, which leads to $VN_1(r_1^{'}) > VN_1(r_1)$ when $r_1^{'} > r_1$. Thus the consistency of $VN_1(r_1)$ in selecting true cotrending rank is established. Analogously, one can establish the consistency of the estimator of the true weak cotrending rank by $VN_2(r_2)$.

(ii) To show the consistency of the joint selection procedure, consider all the possible cases as follows.

Case 1: $r_1^{'} < r_1$.

We have

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = \sqrt{T}\sum_{i=r_1^{'}+1}^{r_1} \widehat{\lambda}_i + O_p(\frac{C_T}{T}),$$

where $\widehat{\lambda}_i$ for $i = r_1^{'} + 1, \cdots, r_1$ is $O_p(1)$ but is not $o_p(1)$.

From Proposition 1 and Lemma 1, the first term dominates, which leads to $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ when $r_1^{'} < r_1$.

Case 2: $r_1^{'} > r_1$.

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = -\sqrt{T} \sum_{i=r_1+1}^{r_1^{'}} \widehat{\lambda}_i + (f(r_1^{'}) - f(r_1))\frac{C_T}{T} + O_p(\frac{C_T}{T^2}),$$

where $\widehat{\lambda}_i$ is $O_p(T^{-1})$ for $i = r_1 + 1, \cdots, m$.

The dominant term in the above equation is $(f(r_1^{'}) - f(r_1))\frac{C_T}{T}$, provided that $\frac{C_T}{\sqrt{T}} \to \infty$, the inequality $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$, holds in this case.

Case 3: $r_1^{'} = r_1$.

When $r_2^{'} > r_2$,

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = -\sqrt{T} \sum_{i=r_2+1}^{r_2^{'}} \widehat{\lambda}_i + (f(r_2^{'}) - f(r_1))\frac{C_T^{'}}{T^2},$$

where $\widehat{\lambda}_i$ is $O_p(T^{-2})$ for $r_2^{'} + 1, \cdots, m$.

Then, we have

$$T^2 \left(VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2)\right) = -\sqrt{T} \sum_{i=r_2+1}^{r_2^{'}} T^2\widehat{\lambda}_i + (f(r_2^{'}) - f(r_2))C_T^{'}.$$

Provided that $\frac{C_T}{\sqrt{T}} \to \infty$, the dominant term is $(f(r_2^{'}) - f(r_2))C_T^{'}$, which is greater than zero. Hence $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ in this case.

When $r_2^{'} < r_2$,

$$VN(r_1^{'}, r_2^{'}) - VN(r_1, r_2) = \sqrt{T} \sum_{i=r_2^{'}+1}^{r_2} \widehat{\lambda}_i + (f(r_2^{'}) - f(r_2))\frac{C_T^{'}}{T^2}.$$

The first term on the right hand side is $O_p(T^{-3/2})$ but is not $o_p(T^{-3/2})$, and dominates the second term, provided that $\frac{C_T^{'}}{T} \to 0$. Hence, $VN(r_1^{'}, r_2^{'}) > VN(r_1, r_2)$ in this case.

Combining the conditions on C_T and $C_T^{'}$ for all the preceding cases, it follows that the joint selection procedure will lead to consistent estimation of the cotrending and weak cotrending ranks.

References

[1] Bacon, D. W., Watts, D. G., 1971. Estimating the transition between two intersecting straight lines. *Biometrika* 58, 525–534.

[2] Bae, Y., Kakkar, V., Ogaki, M., 2006. Money demand in Japan and nonlinear cointegration. *Journal of Money, Credit, and Banking* 38, 1659-1667.

[3] Bai, J., Ng, S., 2002. Determine the number of factors in approximate factor models. *Econometrica* 70, 191-221.

[4] Bai, J., Ng, S., 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72, 1127-1177.

[5] Bierens, H. J., 2000. Nonparametric nonlinear co-trending analysis, with an application to inflation and interest in the U.S.. *Journal of Business and Economic Statistics* 18, 323–337.

[6] Cheng, X., Phillips, P. C. B., 2009. Semiparametric cointegrating rank selection. *Econometrics Journal* 12, S83–S104.

[7] Clements, M. P., Hendry, D. F., 1999. On winning forecasting competitions in economics. *Spanish Economic Review* 1, 123–160.

[8] Engle, R. F., Granger, C. W. J., 1987. Cointegration and error correction: representation, estimation, and testing. *Econometrica* 55, 251-276.

[9] Engle, R. F., Kozicki, S., 1993. Testing for common features. *Journal of Business & Economic Statistics* 11, 369-380.

[10] Harris, D., Poskitt, D. S., 2004. Determination of cointegrating rank in partially non-stationary process via a generalised von-Neumann criterion. *Econometrics Journal* 7, 191-217.

[11] Hatanaka, M., 2000. How to determine the number of relations among deterministic trends., *The Japanese Economic Review* 51, 349-373.

[12] Hatanaka, M., Yamada, H., 2003. Co-trending: A Statistical System Analysis of Economic Trends, Springer.

[13] Hendry, D. F., Mizon, G. E., 1998. Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics* 23, 267-294.

[14] Johansen, S., 1991. Estimation and hypothesis test of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.

[15] Leybourne, S., Newbold, P., Vougas, D., 1998. Unit roots and smooth transitions. *Journal of Time Series Analysis* 19, 83–97.

[16] Lin, C.-F. J., Teräsvirta, T., 1994. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics* 62, 211-228.

[17] Mills, T., 2003. Modelling Trends and Cycles in Economics Time Series, Palgrave Macmillan.

[18] Nelson, C. R., Plosser C. R., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* 10, 139-162.

[19] Ogaki, M., Park, J. Y., 1998. A cointegration approach to estimating preference parameters. *Journal of Econometrics* 82, 107–134.

[20] Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. *Econometrica* 57, 1361-1401.

[21] Perron, P., 1997. Further evidence from breaking trend functions in macroeconomic variables. *Journal of Econometrics* 80, 355-385.

[22] Shintani, M., 2001. A simple cointegrating rank test without vector autoregression. *Journal of Econometrics* 105, 337-362.

[23] Stock, J. H., Watson, M. W., 1988. Testing for common trends. *Journal of the American Statistical Association* 83, 1097-1107.

Table 1. Two dimensional cointegrating rank selection

	$\mathbf{r_1{=}0}$	$r_1 = 1$	$r_1 = 2$	$r_1 = 0$	$\mathbf{r_1{=}1}$	$r_1 = 2$	$r_1 = 0$	$r_1 = 1$	$\mathbf{r_1{=}2}$
				(i) $T = 50$					
RRR-AIC	**0.46**	0.41	0.13	0.00	**0.78**	0.22	0.02	0.55	**0.43**
RRR-BIC	**0.81**	0.17	0.03	0.00	**0.92**	0.08	0.45	0.45	**0.10**
RRR-HQ	**0.62**	0.30	0.07	0.00	**0.85**	0.15	0.13	0.61	**0.26**
VN-AIC	**0.97**	0.03	0.00	0.04	**0.96**	0.00	0.01	0.80	**0.19**
VN-BIC	**1.00**	0.00	0.00	0.65	**0.35**	0.00	0.47	0.52	**0.01**
VN-HQ	**0.99**	0.01	0.00	0.21	**0.79**	0.00	0.09	0.84	**0.07**
				(ii) $T = 100$					
RRR-AIC	**0.49**	0.39	0.12	0.00	**0.78**	0.22	0.00	0.25	**0.75**
RRR-BIC	**0.88**	0.11	0.01	0.00	**0.94**	0.06	0.05	0.73	**0.22**
RRR-HQ	**0.70**	0.25	0.05	0.00	**0.87**	0.13	0.00	0.51	**0.49**
VN-AIC	**0.98**	0.02	0.00	0.00	**1.00**	0.00	0.00	0.39	**0.61**
VN-BIC	**1.00**	0.00	0.00	0.05	**0.95**	0.00	0.01	0.95	**0.05**
VN-HQ	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.76	**0.24**
				(iii) $T = 400$					
RRR-AIC	**0.52**	0.37	0.11	0.00	**0.76**	0.24	0.00	0.00	**1.00**
RRR-BIC	**0.95**	0.05	0.00	0.00	**0.96**	0.04	0.00	0.02	**0.98**
RRR-HQ	**0.80**	0.18	0.03	0.00	**0.89**	0.11	0.00	0.00	**1.00**
VN-AIC	**0.97**	0.03	0.00	0.00	**1.00**	0.00	0.00	0.00	**1.00**
VN-BIC	**1.00**	0.00	0.00	0.00	**1.00**	0.00	0.00	0.06	**0.94**
VN-HQ	**1.00**	0.00	0.00	0.00	**1.00**	0.10	0.00	0.00	**1.00**

Note: Numbers are frequencies of selecting each cointegrating rank. Selection of true rank is shown in bold font.

Table 2. Three dimensional cotrending rank selection: $T = 50$

	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	**0.93**	0.06	0.01	0.00	0.00	0.00
Paired BIC-HQ	**0.93**	0.06	0.01	0.00	0.00	0.00
Joint BIC	**0.81**	0.12	0.08	0.00	0.00	0.00
Joint HQ	**0.79**	0.13	0.08	0.00	0.00	0.00
	(2,2)	**(1,2)**	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	0.02	**0.71**	0.18	0.07	0.02	0.00
Paired BIC-HQ	0.02	**0.78**	0.20	0.00	0.00	0.00
Joint BIC	0.01	**0.46**	0.35	0.14	0.04	0.00
Joint HQ	0.01	**0.57**	0.41	0.00	0.00	0.00
	(2,2)	(1,2)	**(0,2)**	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	0.05	**0.79**	0.01	0.16	0.00
Paired BIC-HQ	0.00	0.05	**0.94**	0.00	0.00	0.00
Joint BIC	0.00	0.02	**0.82**	0.01	0.16	0.00
Joint HQ	0.00	0.02	**0.98**	0.00	0.00	0.00
	(2,2)	(1,2)	(0,2)	**(1,1)**	(0,1)	(0,0)
Paired BIC	0.00	0.00	0.00	**0.81**	0.19	0.00
Paired BIC-HQ	0.00	0.00	0.00	**0.81**	0.19	0.00
Joint BIC	0.00	0.00	0.00	**0.62**	0.38	0.00
Joint HQ	0.00	0.00	0.00	**0.60**	0.40	0.00
	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
Paired BIC	0.00	0.00	0.00	0.03	**0.96**	0.01
Paired BIC-HQ	0.00	0.00	0.00	0.03	**0.97**	0.00
Joint BIC	0.00	0.00	0.00	0.01	**0.97**	0.01
Joint HQ	0.00	0.00	0.00	0.01	**0.99**	0.00
	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	**(0,0)**
Paired BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Paired BIC-HQ	0.00	0.00	0.00	0.00	0.98	**0.02**
Joint BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint HQ	0.00	0.00	0.00	0.00	0.98	**0.02**

Note: The first element in the parenthesis denotes cotrending rank, r_1, the second element denotes weak cotrending rank, r_2. Numbers are frequencies of selecting each pair of cotrending ranks. Selection of true rank is shown in bold font.

Table 3. Three dimensional cotrending rank selection: $T = 100$

	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Paired BIC-HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
	(2,2)	**(1,2)**	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	0.02	**0.93**	0.00	0.06	0.00	0.00
Paired BIC-HQ	0.02	**0.98**	0.00	0.00	0.00	0.00
Joint BIC	0.01	**0.82**	0.00	0.16	0.00	0.00
Joint HQ	0.01	**0.99**	0.00	0.00	0.00	0.00
	(2,2)	(1,2)	**(0,2)**	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	0.03	**0.87**	0.00	0.10	0.00
Paired BIC-HQ	0.00	0.03	**0.97**	0.00	0.00	0.00
Joint BIC	0.00	0.01	**0.89**	0.00	0.10	0.00
Joint HQ	0.00	0.02	**0.98**	0.00	0.00	0.00
	(2,2)	(1,2)	(0,2)	**(1,1)**	(0,1)	(0,0)
Paired BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Paired BIC-HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
Paired BIC	0.00	0.00	0.00	0.02	**0.98**	0.01
Paired BIC-HQ	0.00	0.00	0.00	0.02	**0.98**	0.00
Joint BIC	0.00	0.00	0.00	0.01	**0.98**	0.01
Joint HQ	0.00	0.00	0.00	0.01	**0.99**	0.00
	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	**(0,0)**
Paired BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Paired BIC-HQ	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint HQ	0.00	0.00	0.00	0.00	0.00	**1.00**

Note: See note for Table 2.

Table 4. Three dimensional cotrending rank selection: $T = 400$

	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Paired BIC-HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint BIC	**1.00**	0.00	0.00	0.00	0.00	0.00
Joint HQ	**1.00**	0.00	0.00	0.00	0.00	0.00
	(2,2)	**(1,2)**	(0,2)	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	**0.99**	0.00	0.01	0.00	0.00
Paired BIC-HQ	0.00	**1.00**	0.00	0.00	0.00	0.00
Joint BIC	0.00	**0.94**	0.00	0.06	0.00	0.00
Joint HQ	0.00	**1.00**	0.00	0.00	0.00	0.00
	(2,2)	(1,2)	**(0,2)**	(1,1)	(0,1)	(0,0)
Paired BIC	0.00	0.01	**0.98**	0.00	0.02	0.00
Paired BIC-HQ	0.00	0.01	**0.99**	0.00	0.00	0.00
Joint BIC	0.00	0.00	**0.98**	0.00	0.02	0.00
Joint HQ	0.00	0.00	**1.00**	0.00	0.00	0.00
	(2,2)	(1,2)	(0,2)	**(1,1)**	(0,1)	(0,0)
Paired BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Paired BIC-HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint BIC	0.00	0.00	0.00	**1.00**	0.00	0.00
Joint HQ	0.00	0.00	0.00	**1.00**	0.00	0.00
	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
Paired BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Paired BIC-HQ	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint HQ	0.00	0.00	0.00	0.00	**1.00**	0.00
	(2,2)	(1,2)	(0,2)	(1,1)	(0,1)	**(0,0)**
Paired BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Paired BIC-HQ	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint BIC	0.00	0.00	0.00	0.00	0.00	**1.00**
Joint HQ	0.00	0.00	0.00	0.00	0.00	**1.00**

Note: See note for Table 2.

Table 5. Cotrending rank selection with smooth transition trend models: $T = 400$

	(2,2)	(1,2)	(0,2)	(1,1)	**(0,1)**	(0,0)
			(i) $\gamma = 0.001$			
Paired BIC	0.00	0.00	0.91	0.00	**0.09**	0.00
Paired BIC-HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
Joint BIC	0.00	0.00	0.91	0.00	**0.09**	0.00
Joint HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
			(ii) $\gamma = 0.005$			
Paired BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Paired BIC-HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
Joint BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint HQ	0.00	0.00	1.00	0.00	**0.00**	0.00
			(iii) $\gamma = 0.01$			
Paired BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Paired BIC-HQ	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint BIC	0.00	0.00	0.00	0.00	**1.00**	0.00
Joint HQ	0.00	0.00	0.00	0.00	**1.00**	0.00

Note: The first element in the parenthesis denotes cotrending rank, r_1, the second element denotes weak cotrending rank, r_2. The degree of smoothness in a smooth transition trend model is controled by the scaling parameter γ. Numbers are frequencies of selecting each pair of cotrending ranks. Selection of true rank is shown in bold font.

Table 6. Cotrending relationship among money, income, and interest rates

	Model 1		Model 2		Model 3	
	VN	VN-μ	VN	VN-μ	VN	VN-μ
			(i) $M1$			
Paired BIC	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)
Paired BIC-HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
Joint BIC	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)	(0,1)
Joint HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
			(ii) $M2$			
Paired BIC	(0,1)	(0,2)	(0.1)	(0,2)	(0,1)	(0,2)
Paired BIC-HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
Joint BIC	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)
Joint HQ	(0,1)	(0,2)	(0,1)	(0,2)	(0,1)	(0,2)

Note: The sample period is from 1980Q1 to 2010Q4. Pair of numbers are selected (r_1, r_2) where the first element denotes cotrending rank, r_1, and the second element denotes weak cotrending rank r_2. The first column for each model represents the results of von Neumann criteria from raw series (VN) and the second column represents the results from demeaned series (VN-μ).

Figure 1. Segmented linear trend and smooth transition trend

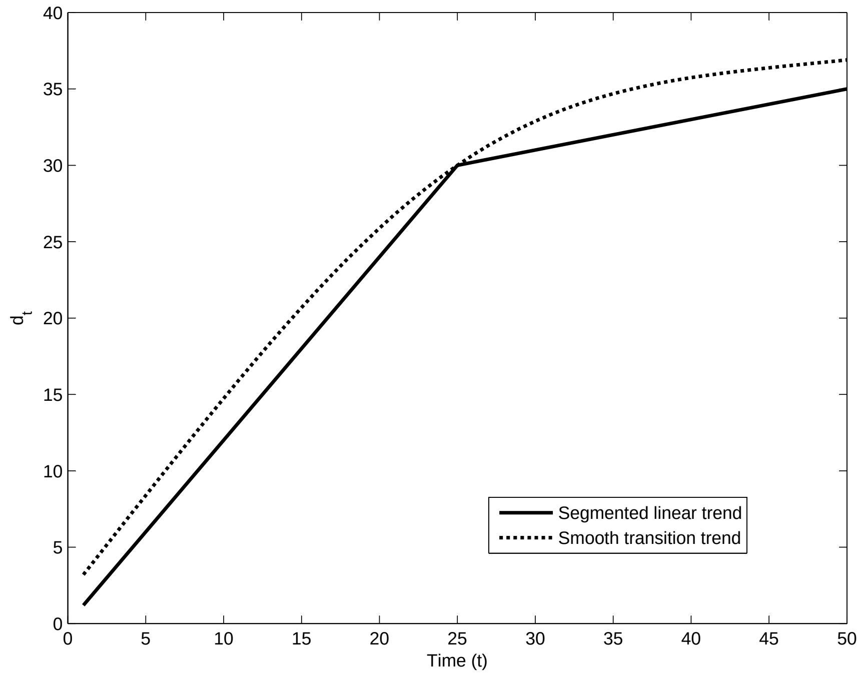

Figure 2. Real GDP

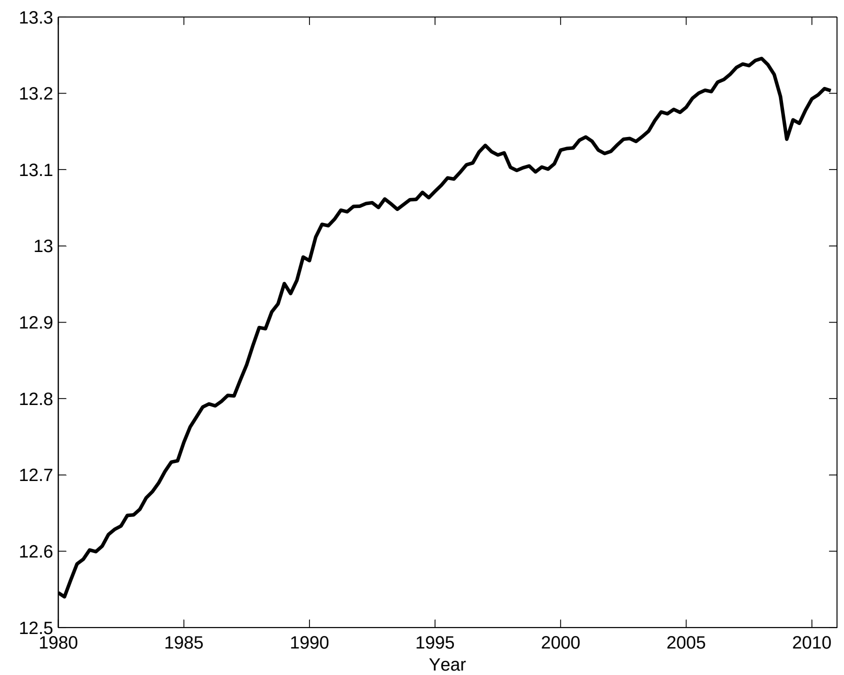

Figure 3. Monetary aggregate: M1

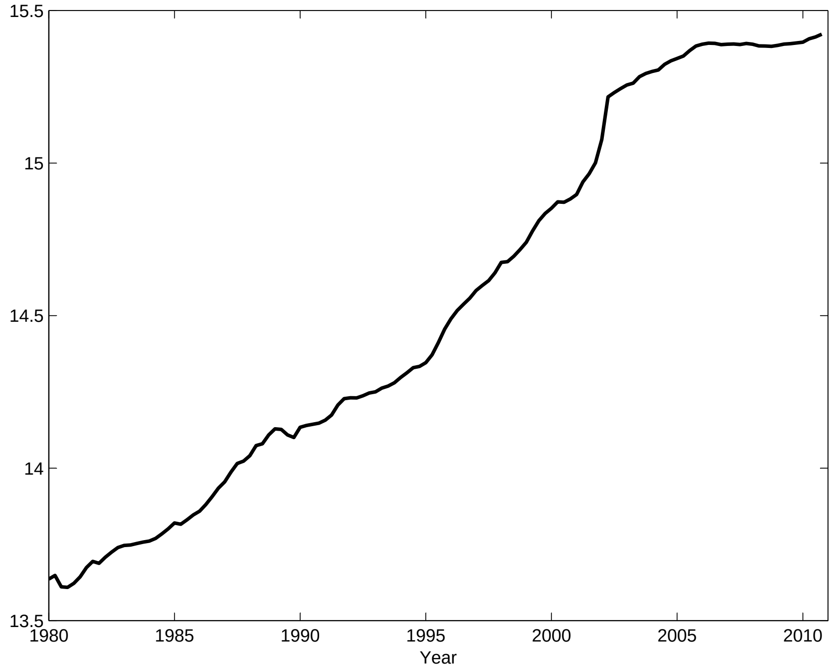

Figure 4. Monetary aggregate: M2

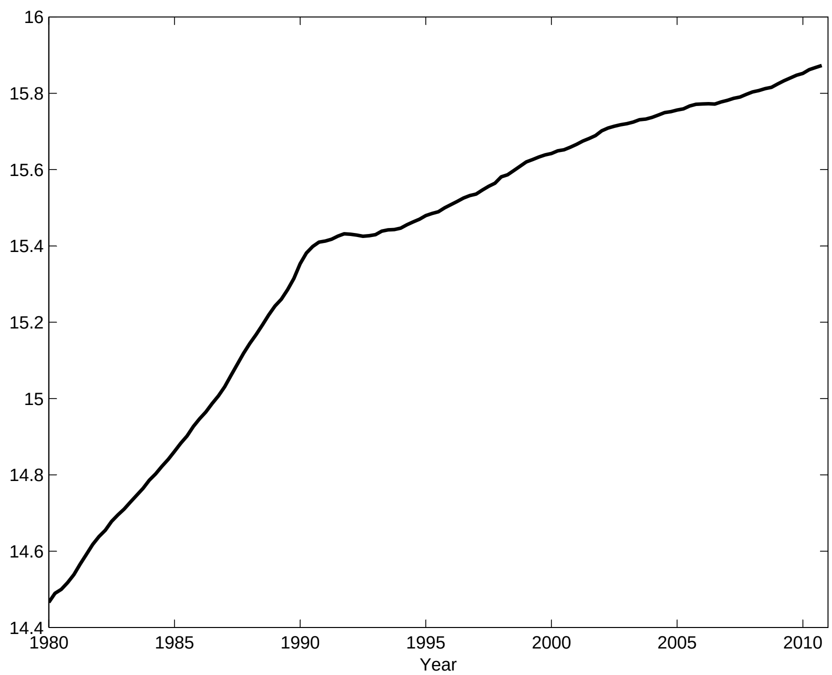

Figure 5. Call rate