Nonparametric Nonlinear Co-Trending Analysis, With an Application to Interest and Inflation in the U.S.

Herman J. Bierens

Pennsylvania State University, Department of Economics, University Park, PA 16802

& Tilburg University, the Netherlands

Given the assumption that the components of a vector time series are stationary about nonlinear deterministic time trends, nonlinear co-trending is the phenomenon that one or more linear combinations of the time series are stationary about a linear trend or a constant, hence the series have common nonlinear deterministic time trends. In this paper we develop nonparametric tests for nonlinear co-trending. These tests are based on generalized eigenvalues, where the two matrices involved are constructed nonparametrically on the basis of partial sums. We apply this approach to the federal funds rate and the CPI inflation rate in the U.S., using monthly data. It appears that these series are nonlinear co-trended, where the nonlinear trend in the inflation rate is positively related to the nonlinear trend in the interest rate. Moreover, the price puzzle seems to a large extent due to this common nonlinear trend. Furthermore, the common nonlinear trend seems to be related to the oil price shocks induced by the OPEC cartel in the early and late seventies.

JEL CODES: C12, C13, C14, C32

KEY WORDS: Nonlinear trends; Trend breaks; Co-trending; Co-breaking; Common trends; Nonparametric tests; Price puzzle.

(*Current version: February 17, 1999*)

1. INTRODUCTION

The aim of this paper is twofold. Our first aim is to develop nonparametric tests for nonlinear co-trending of macroeconomic time series. Given the assumption that the components of a vector time series are stationary about nonlinear deterministic time trends, nonlinear co-trending is the phenomenon that one or more linear combinations of the time series are stationary about a linear trend or a constant, hence the series have common nonlinear deterministic time trends. Second, we want to investigate the nature of the relation between the federal funds rate and the CPI inflation rate in the U.S., in particular whether this relation is due to a common nonlinear deterministic time trend. Since the 1950's these two macroeconomic time series show a remarkable similarity. Moreover, in certain vector autoregressions involving the interest and inflation rate a unit shock in the innovations in the interest rate appears to have a positive effect on the inflation responses. This phenomenon is known as the *price puzzle*. See Bernanke and Blinder (1992), Christiano and Eichenbaum (1992), Christiano, Eichenbaum and Evans (1992, 1995), Eichenbaum (1994), Sims (1995), Balke and Emery (1994, 1995), and Cushman and Zha (1997), among others.

The kind of nonlinear trend stationarity we consider in this paper is $z_t = g(t) + u_t$, where $g(t) = \beta_0 + \beta_1 t + f(t)$, z_t is a k-variate time series process, u_t is a k-variate zero-mean stationary process, and $f(t)$ is a deterministic k-variate nonlinear trend function representing structural change. Nonlinear co-trending is then the phenomenon that there exists a non-zero vector θ such that $\theta^T f(t) = 0$. We shall consider two versions of nonlinear trend stationarity, namely the cases where β_1 is a zero vector or not. The first case applies to our empirical application.

The motivation for considering nonlinear co-trending is fourfold. First, there is now empirical evidence that some long macro-economic time series such as those in the Nelson-Plosser (1982) data set that were initially perceived as unit root processes are probably more in accordance with a nonlinear trend stationary hypothesis. See for example Perron (1988, 1989, 1990) who tested the unit root hypothesis for the Nelson-Plosser series against trend stationarity with a trend break (which is a special case of nonlinear trend stationarity), and Bierens (1997a) who tested the unit root hypothesis for the price level and interest rate series in the extended Nelson-Plosser data set (extended by Schotman and Van Dijk (1991) to 1988) against a smooth nonlinear trend stationarity hypothesis. However, despite the somewhat reluctant conclusion of Bierens (1997a) that the log of the annual

CPI over the period 1860-1988 is probably a nonlinear trend stationary process rather than a unit root process, it appears that for monthly post-war time series the log of the CPI looks more like a unit root process with time varying drift, hence the CPI inflation rate is then a nonlinear trend stationary process.

The second motivation is that quite a few macroeconomic time series that are not unit root processes still behave like cointegrated processes in that the series move together over time in a similar way. But cointegration is only possible for unit root processes, so something else is going on. A possible explanation is that these series have common nonlinear deterministic time trends.

The third motivation is that the (linear trend) stationarity hypothesis as well as the unit root (with constant drift) hypothesis for macroeconomic time series imply that the structure of the economy (i.e. the parameters of the underlying data-generating process) does not change over time. This is quite implausible, in particular for long macro-economic time series such as the Nelson-Plosser (1982) data spanning a century or more.

The fourth motivation is of a philosophical nature. The philosophical question is whether policy interventions should be considered as stochastic events, deterministic events, or a mixture of both. Take for example the federal funds rate (FFR), which is set by the Federal Open Market Committee (FOMC) of the Federal Reserve Board, in order to control inflation. Undoubtly, the FOMC will respond to inflationary signals, and to the extent that these signals are stochastic events, the response of the FOMC will be stochastic as well. However, it is unlikely that the response is completely automatic: It is hard to believe that this discussion is governed by a fixed blueprint. In other words, the actual federal funds rate will reflect both external inflationary signals, which may be considered as stochastic variables, and the subjective assessment of the significance of these stochastic signals by the twelve FOMC members, which may be considered as deterministic and time-varying. It is therefore reasonable to assume that the FFR has a time-varying unconditional expectation.

The general philosophical question is: If a policy maker, or a small body of policy makers, determines the value of a variable Y_t on the basis of information on a vector X_t of variables (possibly including lagged Y_t's), is the conditional expectation of Y_t, given X_t, a time-varying or a time-invariant function of X_t? In the latter case we can write: $Y_t = g(X_t) + U_t$, say, where U_t has conditional expectation equal to zero. But this means that our policy maker has in every situation X_t a "plan" $g(X_t)$

for the action to be taken regarding Y_t, apart from the uncertainty represented by U_t, and that these plans have been the same, and will be the same forever. But policy makers don't live that long, and their tenure will even be shorter. In other words, if, for example, X_t is the federal budget surplus in the US, and Y_t is the way it will be spent, it is not too farfetched to assume that it matters for the outcome of Y_t who the president is, and which party is in control of the Congress. Therefore, it is likely that the function g will change over time, and if so it is possible that the unconditional expectation $E(Y_t)$ will change over time as well. Hamilton (1989) models the occurrence of these regime shifts by a discrete-time, discrete-state first-order Markov process, assuming that the transition probabilities are constant, which eliminates the heterogeneity. However, Hamilton's approach does not solve our philosophical problem; it only shifts it to another level: Are the transition probabilities time-invariant, or not? If not, then $E(Y_t)$ may still be time-varying. In this paper we take the stand that policy interventions by a single policy maker, or a small body of policy makers, will likely have a deterministic component, which may cause heterogeneity of the target variables.

A similar argument applies to the CPI inflation rate as well. In the early seventies and around 1980 the OPEC cartel boosted the producers price of oil[1]. As we will show later, the inflation rate of the producers price index of fuel and related products, and the CPI inflation rate in the US in the period 1968-1994, have strikingly similar camel-back shape patterns (after rescaling), which suggests that the US inflation in this period was to a large extent driven by the oil price shocks. The same applies to the inflation rates in Canada, Japan, Germany, the U.K., and France in this period, although this will not be shown in this paper. In our view the actions of the OPEC were (at least in part) deterministic shocks, in timing as well as in magnitude, since they were undertaken by a relatively small body of policy makers, the OPEC cartel. Therefore, the unconditional expectation of the CPI inflation rate is likely time-dependent, via the oil price shocks. The two oil price shocks, however, had also structural effects. The high oil price made it possible to develop oil fields outside the OPEC areas, such as the North Sea oil fields. Also, they triggered energy conservation in the US and elsewhere. An example is the import, and later the domestic production, of more fuel efficient smaller cars. Thus, the actions of the OPEC also caused structural change in the US economy, and elsewhere.

[1] See URL http://www.york.ac.uk/student/su/essaybank/politics/opec_and_oil_prices.htm for a brief history of the OPEC cartel during the 60-th through the 80-th.

Nonlinear co-trending is of course a special case of a common feature, see Engle and Kozicki (1993). However, the Engle-Kozicki approach requires that the feature involved has to be parametrized. In our case we do not parametrize the nonlinear trend function $f(t)$. Our test is more akin to the KPSS test (Kwiatkowski et al. 1992) for testing the null hypothesis of (trend) stationarity of a univariate time series against the unit root hypothesis.

The plan of the paper is as follows: In Section 2 we discuss the properties of the nonlinear function $g(t)$. In Section 3 we summarize the procedure for testing the number of co-trending vectors and the ideas behind it. In particular, we show how to construct nonparametrically two matrices $\hat{M}_1$ and $\hat{M}_2$ such that their generalized eigenvalues can be used to test for nonlinear co-trending. In Section 4 we derive the asymptotic properties of the matrices $\hat{M}_1$ and $\hat{M}_2$. In Section 5 we derive the actual tests for the number of co-trending vectors on the basis of the generalized eigenvalues of the matrices $\hat{M}_1$ and $\hat{M}_2$, and the asymptotic null distributions of the tests. Also, we propose a test of linear restrictions on the co-trending vectors. In Section 6 we propose consistent estimators of the co-trending vectors. In Section 7 we show what happens if our tests are applied to a cointegrated unit root process rather than a nonlinear trend stationary process. In Section 8 we apply our approach to monthly time series of the federal funds rate and the CPI inflation rate in the U.S. In Section 9 we analyze the consequence of our empirical findings for the price puzzle. In Section 10 we determine the source of the common trend in the CPI inflation rate and the federal funds rate.

The empirical applications involved have been conducted, and can be replicated, by using the author's software package *EasyReg*[2]. The details of the proofs of the technical results in this paper are given in the a separate appendix.[3]

---

[2] *EasyReg* is downloadable from web page
http://econ.la.psu.edu/~hbierens/EASYREG.HTM,
and the data involved can be retrieved from the *EasyReg* database.

[3] This separate appendix is downloadable as an Adobe PDF file from web page
http://econ.la.psu.edu/~hbierens/PAPERS.HTM.

2. NONLINEAR DETERMINISTIC TRENDS AND THEIR PARTIAL SUMS

2.1 Level shifts

Our first example of a nonlinear time trend is the case where $g(t)$ is constant, except for a single jump at time t_1: $g(t) = a$ if $t \le t_1$, $g(t) = b$ if $t > t_1$, where a and b are vectors. If t_1 is constant, it is impossible to asymptotically distinguish $g(t)$ from b, because then $g(t) = a$ only for a finite number of observations. Therefore, in order to make asymptotics work, one has to assume that t_1 is proportional to the number of observations n: $t_1 = \tau n$, where $\tau \in (0,1)$ is fixed. This assumption is common in the literature on testing for a unit root against stationarity around a breaking trend (see for example Perron 1988, 1989, 1990), and the literature on testing for structural change.

We can write $g(t) = c_n + f_n(t)$ for $t = 1,..,n$, where $c_n = b + (a\text{-}b)[\tau n]/n$ is the average of the $g(t)$'s for $t = 1,..n$, and $f_n(t) = (a\text{-}b)(1\text{-}[\tau n]/n)$ if $t \le [\tau n]$, $f_n(t) = \text{-}(a\text{-}b)[\tau n]/n$ if $t > [\tau n]$. Here and in the sequel, $[\tau n]$ denotes the largest integer $\le \tau n$. The function $f_n(t)$ has a number of properties that are fundamental to our approach. Denote for non-negative numbers p,

$$(1) \qquad F_{p,n}(x) = (1/n)\sum_{t=1}^{[nx]}\frac{f_n(t)}{n^p} \quad if\ x \in [n^{-1},1], \quad F_{p,n}(x) = 0 \quad if\ x \in [0,n^{-1}),$$

and let $F(x) = (a-b)(1-\tau)x$ if $x \in [0,\tau]$, $F(x) = (a-b)\tau(1-x)$ if $x \in (\tau,1]$. The role of the number p will become clear in the next subsections. In the case under review, the value $p = 0$ applies. Then

$$(2) \qquad \int\|F_{p,n}(x)\|^2dx = O(1), \quad \int\|F(x)\|^2dx < \infty, \quad \int\|F_{p,n}(x) - F(x)\|^2dx = o(1),$$
$$F_{p,n}(0) = F_{p,n}(1) = 0, \quad F(0) = F(1) = 0,$$

where the integrals here and in the sequel are taken over the unit interval, unless otherwise indicated. Hence, for $p = 0$,

$$(3) \qquad M_{1,p,n} = \int F_{p,n}(x)F_{p,n}(x)^Tdx \to M_1 = \int F(x)F(x)^Tdx.$$

Next, let m be a subsequence of the natural numbers such that $m \to \infty$, $m = o(n)$. Denote for $p \ge 0$,

(4) $$F'_{p,n}(x) = \frac{1}{m}\sum_{j=0}^{m-1}\frac{f_n([nx]+1-j)}{n^p} \quad if\ [nx] \geq m-1, \quad F'_{p,n}(x) = 0 \quad if\ [nx] < m-1,$$

and let $F'(x) = (a-b)(1-\tau)$ if $x \in [0,\tau]$, $F'(x) = -(a-b)\tau$ if $x \in (\tau,1]$. Then for $p = 0$,

(5) $$\int\|F'_{p,n}(x)\|^2dx = O(1), \quad \int\|F'(x)\|^2dx < \infty, \quad \int\|F'_{p,n}(x) - F'(x)\|^2dx = o(1),$$

and consequently,

(6) $$M_{2,p,n} = \int F'_{p,n}(x)F'_{p,n}(x)dx \to M_2 = \int F'(x)F'(x)^Tdx.$$

Note that the matrix $M_{2,p,n}$ can also be written as:

$$M_{2,p,n} = n^{-p-1}\sum_{t=m}^{n}\left((1/m)\sum_{j=0}^{m-1}f_n(t-j)\right)\left((1/m)\sum_{j=0}^{m-1}f_n(t-j)\right)^T.$$

In the case of co-trending, there exists a vector θ such that $\theta^Tg(t) = \theta^Tc_n + \theta^Tf_n(t)$ is constant, hence $\theta^Tf_n(t) = 0$ for $t = 1,..,n$. The latter follows from the fact that c_n is the average of the $g(t)$'s, and that therefore the average of the $f_n(t)$'s is zero. Therefore, this co-trending vector θ is an eigenvector of both matrices $M_{1,p,n}$ and $M_{2,p,n}$, corresponding to a zero eigenvalue, and due to the convergence results (3) and (6), the same applies to the limit matrices M_1 and M_2.

Along the same lines it is easy to show that these results carry over to level shift trend functions $g(t)$ with any finite number of jumps.

2.2 Kinked linear trends

The next example is the case where $g(t)$ is piecewise linear in t, with connected adjacent pieces. Again we assume that the kinks occur at times proportional to n.

For $t = 1,....,n$ we can always write $g(t) = \beta_{0,n} + \beta_{1,n}t + f_n(t)$, where f_n(t) is such that

(7) $$\sum_{t=1}^{n}f_n(t) = 0, \quad \sum_{t=1}^{n}t\,f_n(t) = 0.$$

Now assume for convenience that $\beta_{0,n} = \beta_0$, $\beta_{1,n} = \beta_1$. In order for (7) to hold, there should be at least two kinks in $f_n(t)$, say:

$$f_n(t) = a_1 t \quad for\ t \le \tau_1 n,$$

$$f_n(t) = a_1\tau_1 n + a_{2,n}\left(t - \tau_1 n\right) \quad for\ \tau_1 n < t \le \tau_2 n,$$

$$f_n(t) = a_1\tau_1 n + a_{2,n}(\tau_2 - \tau_1)n + a_{3,n}\left(t - \tau_2 n\right) \quad for\ t > \tau_2 n,$$

where $0 < \tau_1 < \tau_2 < 1$, τ_1, τ_2, and a_1 are fixed, and $a_{2,n}$ and $a_{3,n}$ are such that the two conditions in (7) hold. The latter implies that $a_{2,n}$ and $a_{3,n}$ are proportional to a_1: $a_{i,n} = \zeta_{i-1,n}a_1$, $i = 2,3$, where the $\zeta_{i-1,n}$'s are functions of τ_1,τ_2, and converge to limits ζ_{i-1}, respectively. It follows now easily that, uniformly on $[0,1]$,

$$f_n(nx)/n \to a_1\psi(x) \ \ where \ \ \psi(x) = \begin{cases} x & if \ \ 0 \le x \le \tau_1, \\ \tau_1 + \zeta_1(x-\tau_1) & if \ \ \tau_1 < x \le \tau_2, \\ \tau_1 + \zeta_1(\tau_2-\tau_1) + \zeta_2(x-\tau_2) & if \ \ \tau_2 < x \le 1. \end{cases}$$

Consequently, defining $F(x) = a_1\int_0^x \psi(y)dy$, $F'(x) = a_1\psi(x)$, the results (2), (3), (5), and (6) go through for $p = 1$. Moreover, due to the second part of (7), we now also have that

$$\int F_{p,n}(x)dx = \int F(x)dx = 0. \tag{8}$$

It is easy to verify that these results carry over to piecewise linear trends with more than two kinks, and to broken linear trends where the pieces are no longer connected. Moreover, the assumption $\beta_{0,n} = \beta_0$, $\beta_{1,n} = \beta_1$ is not essential for our argument.

The essential difference between this case and the previous one is the value of p: $p = 1$ versus $p = 0$. The reason for introducing this variable p in the definitions (1) and (4) is therefore to accommodate various types of nonlinear trends.

2.3 Smooth nonlinear trends

The last example is the case where $f_n(t)$ is piecewise quadratic and differentiable. First, let $f_n(t)$ be piecewise quadratic with three pieces:

$$f_n(t) = a_1 t + (b_1/n)t^2 \quad for\ t \le \tau_1 n,$$

$$f_n(t) = a_1\tau_1 n + b_1\tau^2 n + a_{2,n}\left(t - \tau_1 n\right) + (b_{2,n}/n)\left(t - \tau_1 n\right)^2 \quad for\ \tau_1 n < t \le \tau_2 n,$$

$$f_n(t) = a_1\tau_1 n + b_1\tau^2 n + a_{2,n}(\tau_2 - \tau_1)n + b_{2,n}\left(\tau_2 - \tau_1\right)^2 n + a_{3,n}\left(t - \tau_2 n\right)$$

$$+ (b_{3,n}/n)\left(t - \tau_2 n\right)^2 \quad for\ t > \tau_2 n,$$

where τ_1 and τ_2 are as before, and a_1 and b_1 are constant vectors. The differentiability condition, together with condition (7), impose four linear restrictions on the parameters. Therefore, $a_{2,\mathrm{n}}$, $a_{3,\mathrm{n}}$, $b_{2,\mathrm{n}}$, and $b_{3,\mathrm{n}}$ are linear combinations of a_1 and b_1, with coefficients depending on τ_1 and τ_2, and n. These coefficients converge. Similar to the piecewise linear trend case it follows now that there exist two differentiable piecewise quadratic functions $\phi(x)$ and $\psi(x)$ on [0,1] such that $f_n(nx)/n \to a_1\phi(x) + b_1\psi(x)$, uniformly on [0,1].

The case with q quadratic pieces is similar. The differentiability condition then imposes q-1 linear restrictions on the corresponding components of the 2q parameter vectors. Since condition (7) imposes two additional restrictions, q+1 parameter vectors are linear combinations of the remaining q-1 free parameter vectors. Thus, there exist vectors a_j and piecewise quadratic differentiable functions $\phi_j(x)$ on [0,1] such that

$$\frac{f_n([nx]/n)}{n} \to \sum_{j=1}^{q-1} a_j\phi_j(x),$$

uniformly on [0,1]. Defining

$$F'(x) = \sum_{j=1}^{q-1} a_j\phi_j(x), \qquad F(x) = \sum_{j=1}^{q-1} a_j\int_0^x \phi_j(y)dy,$$

the results (2), (3), (5), and (6) go through again, for p = 1.

Note that, with $A = (a_1,..,a_{q-1})$, $\phi(x) = (\phi_1(x),\ldots,\phi_{q-1}(x))^T$, and $\Phi(x) = \int_0^x \phi(y)dy$,

$$M_1 = A\left(\int \Phi(x)\Phi(x)^2 dx\right)A^T, \quad M_2 = A\left(\int \phi(x)\phi(x)^2 dx\right)A^T,$$

Thus in this case we have nonlinear co-trending if the rows of A are linear dependent.

2.4 Assumptions about the nonlinear trend

In view of the arguments in the previous subsections, it is now reasonable to assume that in general:

Assumption 1: *There exists a non-negative number p such that, with* $F_{p,n}(x)$ *and* $F'_{p,n}(x)$ *defined by* (1) *and* (4), *respectively, the conditions* (2) *and* (5) *hold, hence so do* (3) *and* (6). *Moreover, the matrices* M_1, $M_{1,p,n}$, M_2, *and* $M_{2,p,n}$ *in* (3) *and* (6), *respectively, have the same rank.*

In general the number p is unknown, but as will appear, our main results are invariant for p (the only exception is Theorem 5). The last part of Assumption 1 implies that the convergence results (3) and (6) preserve the rank of the matrices involved, and that the eigenvectors corresponding to the zero eigenvalues of these four matrices are the same, in the sense that these eigenvectors span the same subspace of $\mathbb{R}^k$.

3. INTRODUCTION TO NONLINEAR CO-TRENDING ANALYSIS

Consider a k-variate time series process $z_t = g(t) + u_t$, where $g(t) = E(z_t)$ is a nonlinear trend function and u_t is a zero mean stationary process. In this paper we shall design a test of the null hypothesis that there exists a nonzero k-vector θ such that $\theta^T g(t)$ is linear in t (case 1), or a constant (case 2). In other words, we test the null hypothesis that the time series z_t is nonlinear co-trended. Although only case 2 applies to our empirical application, for the sake of generality we shall treat case 1 first, and then show how the results change if case 2 applies.

First, we shall design a test of the null hypothesis $\mathfrak{C}(1)$ that the space of all such co-trending vectors θ has dimension 1, against the alternative $\mathfrak{C}(0)$ that this dimension is zero, i.e., we test $\mathfrak{C}(1)$ against the alternative hypothesis that the only vector θ for which $\theta^T g(t)$ is linear in t is the zero vector. Subsequently, we shall extend our approach to testing $\mathfrak{C}(r)$ against $\mathfrak{C}(s)$ with $0 \leq s < r$, for $r = 1,..,k$.

If $\theta^T g(t)$ is linear in t, then so is $\theta^T f_n(t)$, where $f_n(t)$ is the OLS residual of the regression of $g(t)$ on an intercept and time t, for $t = 1,...,n$. As we have seen, the two conditions in (7) imply that $\theta^T f_n(t) = 0$ for $t = 1,..,n$. This argument, of course, goes both ways: if $\theta^T f_n(t) = 0$ for $t = 1,..,n$, then $\theta^T g(t)$

is linear in t, for $t \leq n$. In either case, θ is the common eigenvector of the matrices M_1 and M_2 corresponding to a zero eigenvalue. However, it is possible that for such a common eigenvector, $\theta^T g(t)$ is not exactly linear in t. An obvious example is where $\theta^T g(t)$ is linear in all but a finite number of t's. It is even possible that for a subsequence t_j, $\theta^T g(t_j)$ is nonlinear, whereas θ is the common eigenvector of the matrices M_1 and M_2 corresponding to a zero eigenvalue, namely if the number of t_j's which are less or equal to n is of order $o(n)$. In any case, however, the exceptional t's are thinly spread over the set of natural numbers. Therefore, the co-trending vectors we are interested in are the common eigenvectors of the matrices M_1 and M_2 corresponding to the zero eigenvalues.

Let

$$\hat{M}_1 = (1/n)\sum_{t=1}^{n} \hat{F}(t/n)\hat{F}(t/n)^T,$$

where

$$\hat{F}(x) = (1/n)\sum_{t=1}^{[nx]}\left(z_t - \hat{\beta}_0 - \hat{\beta}_1 t\right) \ if\ x \in [n^{-1},1], \quad \hat{F}(x) = 0 \ if\ x \in [0,n^{-1}),$$

with $\hat{\beta}_0$ and $\hat{\beta}_1$ the OLS estimates of the vectors of intercepts and slope parameters in the regression of z_t on time t for $t = 1,...,n$. Note that, since $\hat{F}(x)$ is a step function, $\hat{F}(1) = 0$ because of the first condition in (7), and $\hat{F}(0) = 0$ by definition, we can write

$$\hat{M}_1 = \int \hat{F}(x)\hat{F}(x)^T dx.$$

It will be shown that under Assumption 1, the nonlinear co-trending hypothesis $\mathfrak{C}(1)$, and the assumption that u_t is a linear process:

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t-j}, \ where\ \varepsilon_t \sim i.i.d.\ (0,I_k), \tag{9}$$

$n^{-2p}\hat{M}_1 \to M_1$ in probability, and in particular that $n\theta^T\hat{M}_1\theta$ converges in distribution to a functional of a standard Wiener process, times $\theta^T C(1)C(1)^T\theta$.

The factor $\theta^T C(1)C(1)^T\theta$ is a nuisance parameter which we want to get rid of. This can be done by using a Newey-West (1987) type estimator for $\theta^T C(1)C(1)^T\theta$, along the lines in Bierens (1994, p.197), as follows. Let m be a sequence of natural numbers converging to infinity with n at rate $o(n)$, say

(10) $$m = [n^{\alpha}],\ where\ 0 < \alpha < 1,$$

and consider the matrix

$$\hat{M}_2 = \frac{1}{n}\sum_{j=m}^{n}\left((1/m)\sum_{j=0}^{m-1}(z_{t-j} - \hat{\beta}_0 - \hat{\beta}_1(t-j))\right)\left((1/m)\sum_{j=0}^{m-1}(z_{t-j} - \hat{\beta}_0 - \hat{\beta}_1(t-j))\right)^T$$

Note that this matrix can also be written as an integral:

$$\hat{M}_2 = \int \hat{F}^{\prime}(x)\hat{F}^{\prime}(x)^T dx,$$

where

$$\hat{F}^{\prime}(x) = \frac{1}{m}\sum_{j=0}^{m-1}\left(z_{[nx]+1-j} - \hat{\beta}_0 - \hat{\beta}_1([nx]+1-j)\right) \quad if\ [nx] \geq m-1, \quad \hat{F}^{\prime}(x) = 0 \quad if\ [nx] < m-1.$$

It will be shown that $n^{-2p}\hat{M}_2 \to M_2$, and under $\mathfrak{C}(1)$, $n^{\alpha}\theta^T\hat{M}_2\theta \to \theta^T C(1)C(1)^T\theta$ in probability. A test of $\mathfrak{C}(1)$ against $\mathfrak{C}(0)$ can now be based on the minimum solution $\hat{\lambda}_1$, say, of the generalized eigenvalue problem det$(\hat{M}_1 - \hat{M}_2) = 0$; i.e., the test statistic involved is $n^{1-\alpha}\hat{\lambda}_1$. The reason for using this generalized eigenvalue approach is that then asymptotically $\theta^T C(1)C(1)^T\theta$ will cancel out. We shall extend this test to the case of multiple nonlinear co-trending, and to testing linear restrictions on the co-trending vectors θ.

The asymptotic power of these tests depends on the choice of α: the smaller α, the higher the asymptotic power. However, the value $\alpha = 1/2$ appears to be optimal for the convergence of $\hat{M}_2$ to M_2, hence too small an α may cause size distortion. In the empirical application we shall therefore choose $\alpha = 1/2$.

Also, we show that under the hypothesis $\mathfrak{C}(r)$ with $r \geq 1$ the eigenvectors of the matrix $\hat{M}_1$, and the generalized eigenvectors of $\hat{M}_1$ w.r.t. $\hat{M}_2$, corresponding to the r smallest eigenvalues, are $\sqrt{n}$-consistent estimators of the co-trending vectors θ.

4. THE ASYMPTOTIC PROPERTIES OF THE MATRICES $\hat{M}_1$ AND $\hat{M}_2$

Given the linear representation (9), we can always write

(11) $$u_t = C(1)\varepsilon_t + (1-L)D(L)\varepsilon_t = C(1)\varepsilon_t + v_t - v_{t-1},$$

say, where $D(L) = [C(L)-C(1)]/(1-L)$ and $v_t = D(L)\varepsilon_t$. A sufficient condition for the stationarity of u_t and v_t is that:

Assumption 2: *The process u_t has the linear representation* (9), *where the ε_t's are i.i.d. $N_k(0,I)$, and $C(L) = C_1(L)^{-1}C_2(L)$, where $C_1(L)$ and $C_2(L)$ are matrix-valued finite-order lag polynomials, such that all the roots of* $\det(C_1(L))$ *lie outside the unit circle.*

Cf. Engle (1987). This assumption is more restrictive than necessary, but it will keep the argument below transparent, and focussed on the main issues. See Phillips and Solo (1992) for weaker conditions in the case of linear processes. Also, we could assume instead of Assumption 2 that u_t is stationary and ergodic, so that we still can write $u_t = C(1)\varepsilon_t + v_t - v_{t-1}$, where now ε_t is a martingale difference process with unit variance matrix and v_t is a stationary process. Cf. Hall and Heyde (1980, p.136). Note that Assumption 2 does not restrict the lag polynomial $C_2(L)$. However, we do need the additional condition that

Assumption 3: *The matrix $C(1)$ is nonsingular.*

This separation of conditions will prove convenient when we compare nonlinear co-trending with cointegration, in Section 7.

It is a standard exercise in Wiener measure calculus to verify from Assumptions 1-2, the decomposition (11), and the functional central limit theorem, that

(12) $$\sqrt{n}(\hat{F}(x) - n^{p}F_{p,n}(x)) \Rightarrow C(1)\{W_k(x) - xW_k(1) + 3(x^2 - x)[2\int W_k(y)dy - W_k(1)]\} = C(1)\overline{W}_k(x),$$

say, where W_k is a k-variate standard Wiener process, and "$\Rightarrow$" denotes weak convergence. Cf. Billingsley (1968). The random function $\overline{W}_k(x)$ is known as a k-variate detrended standard Wiener process. Moreover, it is not hard to verify, using the decomposition (11), that under Assumptions 1-2,

$$\hat{F}_n^{\prime}(t/n) - n^{p}F_{p,n}^{\prime}(t/n) = C(1)(1/m)\sum_{j=0}^{m-1}\varepsilon_{t+1-j} + \frac{v_{t+1}-v_{t+1-m}}{m} + O_p\left(1/\sqrt{n}\right), \tag{13}$$

for $t = m$-1,..,n-1, where the O_p term is uniform in t. The results (12) and (13) imply:

Lemma 1: *Let Assumptions 1-2 and the hypothesis* $\mathfrak{C}(1)$ *be true, let* θ *be the normalized eigenvector of the matrix* M_1 *corresponding to the eigenvalue* $\lambda_1 = 0$, *and let* $Q = (\theta, Q_*)$, *where* Q_* *is the matrix of orthonormal eigenvectors corresponding to the positive eigenvalues* $\lambda_2 \leq \ldots \leq \lambda_k$ *of* M_1. *Denote* $\Lambda_* = \mathrm{diag}(\lambda_2,..,\lambda_k)$. *Then*:

$$\begin{aligned}\hat{\Omega}_1 &= \begin{pmatrix}\sqrt{n} & 0^T \\ 0 & n^{-p}I_{k-1}\end{pmatrix} Q^T\hat{M}_1 Q \begin{pmatrix}\sqrt{n} & 0^T \\ 0 & n^{-p}I_{k-1}\end{pmatrix} \\ \rightarrow \Omega_1 &= \begin{pmatrix}\theta^T C(1)\int\overline{W}_k(x)\overline{W}_k(x)^T dx C(1)^T\theta & \theta^T C(1)^T\int\overline{W}_k(x)F(x)^T dx Q_* \\ Q_*^T\int F(x)\overline{W}_k(x)^T dx C(1)^T\theta & \Lambda_*\end{pmatrix}\end{aligned} \tag{14}$$

in distribution, and for $m \rightarrow \infty$ *at rate* $o(n)$,

$$\begin{aligned}\hat{\Omega}_2 &= \begin{pmatrix}\sqrt{m} & 0^T \\ 0 & n^{-p}I_{k-1}\end{pmatrix} Q^T\hat{M}_2 Q \begin{pmatrix}\sqrt{m} & 0^T \\ 0 & n^{-p}I_{k-1}\end{pmatrix} = \Omega_2 + O_p(\sqrt{m/n}) + O_p(1/\sqrt{m}), \\ &\textit{where } \Omega_2 = \begin{pmatrix}\theta^T C(1)C(1)^T\theta & 0^T \\ 0 & Q_*^T M_2 Q_*\end{pmatrix}.\end{aligned} \tag{15}$$

Note that the rate of convergence to zero of the O_p terms in (15) are optimal for m proportional to $\sqrt{n}$. For this reason we have chosen this m in our empirical application as in (10), with $\alpha = 1/2$.

Clearly, the matrices $n^{-p}\hat{M}_1$ and $n^{-p}\hat{M}_2$ converge to singular matrices M_1 and M_2, respectively. As is well-known, the generalized eigenvalue problem $\det(M_1 - \lambda M_2) = 0$ is ill-defined if M_2 is singular. Since in our case both matrices are singular, the result of Anderson, Brons and Jensen (1983), which is used by Johansen (1988,1991,1994), Johansen and Juselius (1990), and Bierens (1997b) to prove that ordered generalized eigenvalues converge in distribution, does not apply them. However, similarly to Bierens (1997b), it follows straightforwardly from (14) and (15) that we can use their rescaled inverses:

Lemma 2: *Let Assumptions 1-3 and the hypothesis* $\mathfrak{C}$(1) *hold. For every subsequence* $m = o(n)$ *of natural numbers we have*:

$$\hat{\Delta}_1 = \begin{pmatrix} 1/\sqrt{m} & 0^T \\ 0 & n^{p} I_{k-1} \end{pmatrix} Q^T \left(\frac{m}{n} \hat{M}_1^{-1} \right) Q \begin{pmatrix} 1/\sqrt{m} & 0^T \\ 0 & n^{p} I_{k-1} \end{pmatrix} \to \Delta_1 = \begin{pmatrix} \tilde{\mu}^{-1} & 0^T \\ 0 & O \end{pmatrix}, \tag{16}$$

in distribution, where

$$\begin{aligned} &= \theta^T C(1) \int \overline{W}_k(x) \overline{W}_k(x)^T dx C(1)^T \theta - \theta^T C(1) \int \overline{W}_k(x) F(x)^T dx Q_* \Lambda_*^{-1} Q_*^T \int F(y) \overline{W}_k(y)^T dy C(1)^T \theta \\ &\sim \theta^T C(1) C(1)^T \theta \left(\int (\overline{W}_1(x))^2 dx - \int \overline{W}_1(x) F(x)^T dx Q_* \Lambda_*^{-1} Q_*^T \int F(y) \overline{W}_1(y) dy \right), \end{aligned} \tag{17}$$

and

$$\underset{n\to\infty}{\text{plim}} \hat{\Omega}_2^{-1} = \Omega_2^{-1}. \tag{18}$$

5. THE TESTS

5.1 The generalized eigenvalue problem

Observe that the ordered solutions $\hat{\lambda}_1 \le \hat{\lambda}_2 \le .... \le \hat{\lambda}_k$ of the generalized eigenvalue problem

$$\det\left(\hat{M}_1 - \lambda \hat{M}_2\right) = 0 \tag{19}$$

are related to the decreasingly ordered solutions $\tilde{\lambda}_1 \ge \tilde{\lambda}_2 \ge .... \ge \tilde{\lambda}_k$ of the generalized eigenvalue problem

$$\det(\hat{\Delta}_1 - \lambda \hat{\Omega}_2^{-1}) = 0 \tag{20}$$

by the equality $(n/m)\hat{\lambda}_j = \tilde{\lambda}_j^{-1}$, $j = 1, \dots, k$. Moreover, it follows from Lemma 2 and Anderson, Brons and Jensen (1983) that the ordered solutions of generalized eigenvalue problem (20) converge in distribution to the ordered solutions of the generalized eigenvalue problem

$$\det(\Delta_1 - \lambda\,\Omega_2^{-1}) = 0. \tag{21}$$

Clearly, under the hypothesis $\mathfrak{C}(1)$ all but one of the solutions of (21) are zero. The only non-zero solution is $\lambda_1 = (\theta^T C(1)C(1)^T\theta)\tilde{\mu}^{-1}$, so that by (17),

$$(n/m)\hat{\lambda}_1 \to \int\left(\overline{W}_1(x)\right)^2 dx - \int\overline{W}_1(x)F(x)^T dx Q_*\Lambda_*^{-1}Q_*^T\int F(x)\overline{W}_1(x)dx \tag{22}$$

in distribution.

Similarly, comparing the solutions of generalized eigenvalue problem $\det(\hat{\Omega}_1 - \lambda\hat{\Omega}_2) = 0$ with the solutions of eigenvalue problem (19), it follows easily that under Assumptions 1-3 and the hypothesis $\mathfrak{C}(1)$, the vector $(\hat{\lambda}_2,...,\hat{\lambda}_k)$ converges in probability to the vector of ordered solutions of the generalized eigenvalue problem $\det(\Lambda_* - \lambda Q_*^T M_1 Q_*) = 0$, which are all positive. It is now easy to verify that the following more general result holds:

Theorem 1: *Let Λ_{k-r} be the diagonal matrix of the k-r largest eigenvalues of the matrix M_1, and let Q_{k-r} be the matrix of corresponding orthonormal eigenvectors. Choose $m = [n^{\alpha}]$ with $0 < \alpha < 1$. Under Assumptions 1-3 and null hypothesis $\mathfrak{C}(r)$, the solution $\hat{\lambda}_r$ of increasingly ordered solutions of the generalized eigenvalue problem* (19) *satisfies $n^{1-\alpha}\hat{\lambda}_r \to \overline{\lambda}_r$ in distribution, where $\overline{\lambda}_r$ is the maximum eigenvalue of the matrix*

$$\int\overline{W}_r(x)\overline{W}_r(x)^T dx - \int\overline{W}_r(x)F(x)^T dx\, Q_{k-r}\Lambda_{k-r}^{-1}Q_{k-r}^T\int F(y)\overline{W}_r(y)^T dy.$$

Under the alternative hypothesis $\mathfrak{C}(s)$, with s < r, we have: $\operatorname{plim}_{n\to\infty}\hat{\lambda}_r > 0$.

5.2 The asymptotic size of the test

A practical problem with this test is that the null distribution is case-dependent: it depends on F. In particular, it is easy to verify that the right-hand side of (22) is equal to

$$\inf_{\beta}\int\left(\overline{W}_1(x) - \beta^T F(x)\right)^2 dx = \int\overline{W}_1(x)^2 dx - \tilde{\beta}^T M_1\tilde{\beta}, \quad \textit{where } \tilde{\beta} = M_1^{+}\int F(x)\overline{W}_1(x)dx,$$

with M_1^+ the generalized inverse of M_1. Thus, the limiting null distribution involved is the same as the limiting distribution of n^{-1} times the residual sum of squares of the (spurious) regression of $\bar{W}_1(t/n)$ on $F(t/n)$ for t = 1,..,n. This result suggests, in the case of testing the null hypothesis $\mathfrak{C}(1)$, the following way to simulate the null distribution. Generate a version of $\bar{W}_1(x)$ independently of the data, by drawing n independent standard normal random variables e_j for j = 1,..,n, regressing them on an intercept and j, and forming partial sums of the residuals $\tilde{e}_j$, divided by $\sqrt{n}$:

$$\bar{W}_{1,n}(x) = (1/\sqrt{n})\Sigma_{j=1}^{[nx]}\tilde{e}_j \text{ for } x \in [n^{-1},1], \quad \bar{W}_{1,n}(x) = 0 \text{ for } x \in [0,n^{-1}).$$

Then $\bar{W}_{1,n} \Rightarrow \bar{W}_1^*$ for $n \to \infty$, where the latter has the same distribution as $\bar{W}_1$, but is now independent of the data. Since by Assumption 1 and (12), $n^{-p}\hat{F}(x)$ converges to $F(x)$ in L^2, it follows now that for fixed $K > 0$,

$$\lim_{n\to\infty} P\left(\inf_{\beta}\int(\bar{W}_{n,1}(x) - \beta^T\hat{F}(x))^2 dx > K\right) = P\left(\inf_{\beta}\int(\bar{W}_1(x) - \beta^T F(x))^2 dx > K\right).$$

By replicating this procedure N times, where N goes to infinity with n, we get consistent estimates of the quantiles of the null distribution (22), provided that the rank of the matrix M_1 is indeed k-1. If the latter is not true, we end up with a lowerbound or an upperbound, depending on whether the actual rank of M_1 is k or less than k-1. Since we are testing the null hypothesis that rank(M_1) = k-1 against the alternative that rank(M_1) = k, the simulation procedure under review will therefore have a negative effect on the power of the test.

A much simpler way to get around the problem that the null distribution of our test is case-dependent, is to base the asymptotic critical values of the test of the null hypothesis $\mathfrak{C}(r)$ on the maximum eigenvalue of the matrix

$$\int \bar{W}_r(x)\bar{W}_r(x)^T dx \tag{23}$$

only. A motivation for this choice is that under $\mathfrak{C}(r)$, the function $F(x)$ may be considered as an unknown parameter in a space Ξ_r, where

Definition 1: Ξ_r *is the space of k-dimensional functions F on* [0,1] *for which* $F(0) = F(1) = 0$, $\int F(x)dx = 0$, $\int F(x)^T F(x)dx < \infty$, *and* rank$[\int F(x)F(x)^T dx] = k-r$, and that therefore the null hypothesis of co-trending is a composite hypothesis, similar to the case of testing the composite null hypothesis that a parameter (vector) ω is contained in set Ω_0, with $\zeta(\omega)$ the size of the test in a point ω in Ω_0. As is well-known, the classical approach in this case is to define the actual size of this test by $\sup_{\omega\in\Omega_0}\zeta(\omega)$. The following result shows that if we base the critical value of our test on the maximum eigenvalue of (23), we actually get the correct size, in the classical sense:

Theorem 2: *There exists a sequence F_m in Ξ_r such that*

(24) $$\text{plim}_{m\to\infty}\int\overline{W}_r(x)F_m(x)^T dx Q_*\Lambda_*^{-1}Q_*^T\int F_m(y)\overline{W}_r(y)^T dy = O.$$

Consequently, under Assumptions 1-3 and the null hypothesis $\mathfrak{C}(r)$,

(25) $$\sup_{F\in\Xi_r}\lim_{n\to\infty}P(n^{1-\alpha}\hat{\lambda}_r \geq K) = P(\overline{\lambda}_r^* \geq K)$$

for every $K > 0$, where $\overline{\lambda}_r^*$ is the maximum eigenvalue of the matrix (23).

Therefore, if we interpret $F(x)$ as a nuisance parameter, with parameter space Ξ_r, and K the critical value, the actual asymptotic size of the test is given by (25).

The proof of (24) in the case $r = 1$ is based on Mercer's theorem. See Dunford and Schwartz (1963, p.1088), and Bierens and Ploberger (1997). The details of the proof can be found in the separate appendix. The main idea is to arrange the eigenfunctions $\psi_j(x)$ and corresponding non-negative eigenvalues λ_j of the covariance function $\Gamma(x,y) = E(\overline{W}_1(x)\overline{W}_1(y))$ such that the subsequence of positive eigenvalues forms a non-increasing sequence. Then we may choose $F_m(x) = Q_*\Lambda_*^{1/2}(\psi_{m+1}(x),..,\psi_{m+k-1}(x))^T$, where the $\psi_j(x)$'s correspond to positive eigenvalues.

The 80%, 90% and 95% quantiles of the distribution of the random variable $\overline{\lambda}_r^*$ for $r = 1,..,5$ are given in Table 1. These quantiles are calculated by Monte Carlo simulation, on the basis of 10,000 replications of samples of size $n = 500$ from the $N_r(0,I_r)$ distribution.

Table 1: *Values of K for which* $P(\bar{\lambda}_r^* \le K) = p$

p:	0.80	0.90	0.95
r:		K	
1	0.091103	0.119616	0.150989
2	0.134492	0.169183	0.202642
3	0.173114	0.214069	0.252212
4	0.205922	0.251317	0.294746
5	0.236006	0.282870	0.330943

5.3 Testing linear restrictions on the co-trending vectors

Once we have established the number r of linear independent co-trending vectors θ, we may wish to test the null hypothesis that the columns of a given $k \times s$ matrix H with $1 \le s \le r$ span a subspace of the space of co-trending vectors, similar to testing linear restrictions on the cointegrating vectors by Johansen's (1988,1991,1994) likelihood ratio approach. It is straightforward to verify from Lemma 1:

Theorem 3: *Let Assumptions 1-3 and the hypothesis* $\mathfrak{C}(r)$ *hold, and let* $m = [n^\alpha]$, *with* $0 < \alpha < 1$. *Let H be a given* $k \times s$ *matrix with* $1 \le s \le r$, *and let* $\tilde{\lambda}_s$ *be the maximum solution of the generalized eigenvalue problem* $\det[H^T\hat{M}_1 H - \lambda H^T\hat{M}_2 H] = 0$. *If the columns of H span a subspace of the space of co-trending vectors then* $n^{1-\alpha}\tilde{\lambda}_s$ *converges in distribution to the maximum eigenvalue* $\bar{\lambda}_s^*$ *of the matrix* $\int \bar{W}_s(x)\bar{W}_s(x)^T dx$, *whereas otherwise* $\tilde{\lambda}_s$ *converges in probability to a positive constant.*

5.4 Detrending or not?

All our results so far are based on detrended data. But there are situations where there is no linear trend in the data, for example when we use differenced time series. Then taking out a constant mean, by subtracting from z_t its sample mean $\bar{z}$, will suffice. It is easy to verify that all our results carry over, provided that we replace the process $\bar{W}_k(x)$ defined in (12) by a k-variate standard Brownian bridge $W_k^o(x) = W_k(x) - xW_k(1)$. Moreover, instead of the critical values in Table 1 we should use the ones in Table 2, where λ_r^o is the maximum eigenvalue of the matrix $\int W_r^o(x)W_r^o(x)^T dx$.

Table 2: Values of K for which $P(\lambda_r^o \leq K) = p$			
p:	0.80	0.90	0.95
r:		K	
1	0.2451126	0.3518246	0.4657737
2	0.3993106	0.5356136	0.6742039
3	0.5413243	0.7036614	0.8603746
4	0.6778114	0.8618191	1.0345377
5	0.8170006	1.0141629	1.2194813

6. CONSISTENT ESTIMATION OF THE CO-TRENDING VECTORS

Given the hypothesis $\mathfrak{C}(1)$, there are three candidates for the estimator $\hat{\theta}$, say, of the co-trending vector θ, namely the eigenvector corresponding to the minimum eigenvalue of the matrix $\hat{M}_1$, the minimum eigenvalue of the matrix $\hat{M}_2$, or the minimum generalized eigenvalue of $\hat{M}_1$ w.r.t. $\hat{M}_2$.

First, consider the case where $\hat{\theta}$ is a normalized eigenvector of $\hat{M}_1$. Let $Q = (\theta, Q_*)$ be as in Lemma 1, let $\hat{\xi}$ be the eigenvector of $Q^T\hat{M}_1Q$ corresponding to the minimum eigenvalue $\hat{\lambda}_{1,1}$, say, normalized such that $\hat{\xi}^T = (1, \hat{\xi}_*^T)$, and let $\tilde{\theta} = Q\hat{\xi} = \theta + Q_*\hat{\xi}_*$. Note that $\tilde{\theta}$ is an eigenvector of $\hat{M}_1$ corresponding to the minimum eigenvalue $\hat{\lambda}_{1,1}$. Then

$$Q_*^T\hat{M}_1\tilde{\theta} = Q_*^T\hat{M}_1\theta + Q_*^T\hat{M}_1Q_*\hat{\xi}_* = \hat{\lambda}_{1,1}\hat{\xi}_*,$$

hence

$$n^p\sqrt{n}\hat{\xi}_* = -\left(n^{-2p}Q_*^T\hat{M}_1Q_* - n^{-2p}\hat{\lambda}_{1,1}I_{k-1}\right)^{-1}n^{-p}\sqrt{n}Q_*^T\hat{M}_1\theta .$$

It follows now straightforwardly from Lemma 1 that

$$n^p\sqrt{n}(\tilde{\theta} - \theta) = Q_*n^p\sqrt{n}\hat{\xi}_* \to -Q_*\Lambda_*^{-1}Q_*^T\int F(x)\overline{W}_k(x)^T dxC(1)^T\theta \quad \textit{in distr.}, \tag{26}$$

and consequently, using the trivial equality $\|\tilde{\theta}\| - \|\theta\| = (\tilde{\theta}+\theta)^T(\tilde{\theta}-\theta)/\left(\|\tilde{\theta}\| + \|\theta\|\right)$, and the fact that $\|\theta\| = 1$,

$$(27) \qquad n^{p}\sqrt{n}(\|\tilde{\theta}\| - 1) \to -\theta^T Q_* \Lambda_*^{-1} Q_*^T \int F(x)\overline{W}_k(x)^T dx C(1)^T \theta \quad in\ distr.$$

Combining (26) and (27), and denoting $\hat{\theta} = \tilde{\theta}/\|\tilde{\theta}\|$, now yield:

$$(28) \qquad n^{p}\sqrt{n}(\hat{\theta} - \theta) \to -\left(I_k - \theta\theta^T\right) Q_* \Lambda_*^{-1} Q_*^T \int F(x)\overline{W}_k(x)^T dx C(1)^T \theta \quad in\ distr.$$

Next, let us see what happens under the hypothesis $\mathfrak{C}(1)$ if $\hat{\theta}$ is the normalized eigenvector of the matrix $\hat{M}_2$ corresponding to the minimum eigenvalue. As in the previous case, the limiting distribution and rate of convergence of $\hat{\theta} - \theta$ is crucially determined by part (15) of Lemma 1. It can be verified from the proof of Lemma 1 in the separate appendix that the indicated rate of convergence involved is too conservative, and that it is possible that $n^{-p}\sqrt{n}\theta^T \hat{M}_2 \theta_* = O_p(1)$, provided that $p \geq 1/2$. If so, then under some additional regularity conditions, $\sqrt{n}(\hat{\theta} - \theta)$ converges in distribution. However, if $p < 1/2$ then $|n^{-p}\sqrt{n}\theta^T \hat{M}_2 \theta_*|$ converges to infinity, and so will $\|\sqrt{n}(\hat{\theta} - \theta)\|$. Since in general p is unknown, we therefore recommend against using this estimator.

Finally, let now $\hat{\xi}$ be the generalized eigenvector corresponding to the minimum solution $\hat{\lambda}_1$ of the generalized eigenvalue problem (19), again normalized such that $\hat{\xi}^T = (1, \hat{\xi}_*^T)$, and let $\tilde{\theta}$ be as before. Then

$$Q_*^T \hat{M}_1 \tilde{\theta} - \hat{\lambda}_1 Q_*^T \hat{M}_2 \tilde{\theta} = Q_*^T\left(\hat{M}_1 - \hat{\lambda}_1 \hat{M}_2\right)\theta + Q_*^T\left(\hat{M}_1 - \hat{\lambda}_1 \hat{M}_2\right) Q_* \hat{\xi}_* = 0,$$

hence

$$n^{p}\sqrt{n}\hat{\xi}_* = -\left(n^{-2p} Q_*^T \hat{M}_1 Q_* - n^{-2p}\hat{\lambda}_1 Q_*^T \hat{M}_2 Q_*\right)^{-1}\left(n^{-p}\sqrt{n} Q_*^T \hat{M}_1 \theta - \sqrt{n/m}\hat{\lambda}_1 n^{-p}\sqrt{m} Q_*^T \hat{M}_2 \theta\right)$$

$$= -\left(n^{-2p} Q_*^T \hat{M}_1 Q_*\right)^{-1} n^{-p}\sqrt{n} Q_*^T \hat{M}_1 \theta + o_p(\sqrt{m/n}),$$

where the last equality follows from Lemma 1 and the fact that by Theorem 1, $\hat{\lambda}_1 = O_p(m/n)$. Therefore, the results (26), (27), and (28) carry over.

Along the same lines it can be shown that in general the following result holds:

Theorem 4: *Let Assumptions 1-3 and the hypothesis* $\mathfrak{C}(r)$ *hold. Let* $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_r)$ *and* $\theta = (\theta_1, \dots, \theta_r)$ *be the* $k \times r$ *matrices of either the orthonormal eigenvectors of the matrices* $\hat{M}_1$ *and* M_1

alone, or the orthonormal generalized eigenvectors of $\hat{M}_1$ *w.r.t.* $\hat{M}_2$ *and* M_1 *w.r.t.* M_2*, respectively, corresponding to the r smallest eigenvalues. Let* Q_* *be the matrix of the other k-r orthonormal eigenvectors of* M_1*, and let* Λ_* *be the diagonal matrix of corresponding k-r positive eigenvalues. Then jointly for* $i = 1,..,r$,

$$n^{p}\sqrt{n}(\hat{\theta}_i - \theta_i) \rightarrow -\left(I_k - \theta_i\theta_i^T\right)\left(Q_*\Lambda_*^{-1}Q_*^T\right)\left(\int F(x)\bar{W}_k(x)^T dx\right)C(1)^T\theta_i \text{ in distr.}$$

7. COMPARISON WITH COINTEGRATION

The nonlinear trend stationarity hypothesis and the unit root with drift hypothesis are difficult to distinguish. Therefore, we shall also derive the asymptotic distribution of our test under the unit root hypothesis with possible cointegration. Thus, let the data generating process now be $\Delta z_t = \beta_1 + u_t$, where u_t obeys Assumption 2 (but not Assumption 3 of course), hence $z_t = \beta_0 + \beta_1 t + \sum_{j=1}^{t} u_j$, where $\beta_0 = z_0$. Now denote

$$W_k^*(x) = W_k(x) + (6x-4)\int W_k(y)dy - (12x-6)\int yW_k(y)dy, \quad W_k^{**}(x) = \int_0^x W_k^*(y)dy.$$

Then it can be shown, similarly to Bierens (1997b):

Theorem 6: *Let* $z_t = z_{t-1} + \beta_1 + u_t$ *be a k-variate unit root with drift process with* u_t *obeying Assumption 2, and let* $m = [n^{\alpha}]$ *for some* $\alpha \in (0,1)$. *Suppose there are r cointegrating vectors, and that for each cointegrating vector* θ, $\theta^T D(1)D(1)^T\theta > 0$. *Let* $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq .... \leq \hat{\lambda}_k$ *be the ordered solutions of the generalized eigenvalue problem* (19). *Then the vector* $n^{1-\alpha}(\hat{\lambda}_1,..,\hat{\lambda}_r)$ *converges in distribution to the vector of ordered eigenvalues of the matrix*

$$\int \bar{W}_r(x)\bar{W}_r(x)^T dx - \int \bar{W}_r(x)W_{k-r}^{**}(x)^T dx\left(\int W_{k-r}^{**}(x)W_{k-r}^{**}(x)dx\right)^{-1}\int W_{k-r}^{**}(x)\bar{W}_r(x)^T dx,$$

and $(\hat{\lambda}_{r+1},\ldots,\hat{\lambda}_k)$ *converges in distribution to the vector of ordered solutions of the generalized eigenvalue problem*

$$\det\left[\int W_{k-r}^{**}(x)W_{k-r}^{**}(x)^T dx - \lambda\int W_{k-r}^{*}(x)W_{k-r}^{*}(x)^T dx\right] = 0.$$

If $\beta_1 = 0$*, so that* z_t *is a k-variate unit root process without drift, and if* z_t *is demeaned rather than* *detrended, the above results carry over, provided that* $\overline{W}_r(x)$ *is replaced by the r-variate standard Brownian bridge* $W_r^0(x) = W_r(x) - xW_r(1)$ *and* $W_{k-r}^{**}(x)$ *by* $\tilde{W}_{k-r}^{**}(x) = \int_0^x \tilde{W}_{k-r}^{*}(x)dx$*, where* $\tilde{W}_{k-r}^{*}(x) = W_{k-r}(x) - \int W_{k-r}(y)dy$.

This result shows that our nonlinear co-trending tests are incapable of distinguishing nonlinear co-trending from cointegration. Therefore, before we apply the co-trending tests we should test first whether the series involved are unit root processes.

8. NONLINEAR CO-TRENDING ANALYSIS OF INTEREST AND INFLATION

8.1 The data

We use monthly time series of the federal funds rate (FFR) and the CPI inflation rate (CPIR) (i.e., the annual percentage change of the consumer price index), for months 1954.07 through 1994.12. The two standardized[4] series are plotted in Figure 1.

<Insert Figure 1 about here>

We see clearly that these series have a common pattern. The cross-correlation between FFR and lagged CPIR, and CPIR and lagged FFR, is maximal 0.7747 for lag = 0. The dates 1974.11 and 1980.05 correspond to the top of the two main peaks in the CPIR, which approximately correspond to the two peaks in the FFR as well. Moreover, it is also clear from Figure 1 that the series do not have a linear trend, or, if they are unit root processes, do not have drift. Therefore our nonlinear co-trending tests can be conducted without detrending.

However, since our test becomes a cointegration test if the series are unit root processes, we have first conducted a variety of unit root and stationarity tests. The tests involved are the Phillips-Perron (1988) unit root test (PP), Bierens' (1993) unit root tests HOAC(1,1) and HOAC(2,2) on the basis of higher order sample autocorrelations, the Bierens-Guo (1993) tests 1 through 4, indicated below by BG(1) through BG(4), of the stationarity hypothesis, and the KPSS test of the stationarity hypothesis. Also, we have conducted these tests to the first difference of log(CPI), denoted by

---

[4] Each series $x(t)$ is standardized between 0 and 1 by applying the transformation $y(t) = (x(t) - \min_{1\le t\le n}\{x(t)\})/(\max_{1\le t\le n}\{x(t)\} - \min_{1\le t\le n}\{x(t)\})$

Δlog(CPI), because CPIR is approximately proportional to the 12 months moving average of Δlog(CPI). Hence if the latter time series is stationary then so is CPIR, but nevertheless CPIR will then look like a unit root process.

The test results are mixed[5]. Some of the unit root tests reject the unit root hypothesis, some don't, and some of the stationarity tests reject stationarity, while some don't. In particular, the KPSS test rejects the stationarity hypothesis for all three series. The KPSS test, however, will also have power against nonlinear trend stationarity (and other non-unit root alternatives as well: see for example Lee and Schmidt 1996), because our test for nonlinear co-trending becomes the KPSS test if applied to a single time series.

The mixed test results indicate that these three series are neither genuine unit root processes nor genuine stationary processes. In particular, if one concludes that Δlog(CPI) is not a unit root process, then neither is CPIR, despite the test results for the latter series.

We also have conducted Bierens' (1997a) tests of the unit root hypothesis against nonlinear trend stationarity, but the unit root hypothesis could not be rejected. The latter might be due to the lack of smoothness of the nonlinear trends. Despite these results, the general conclusion from the other tests is that the FFR and the CPIR are neither genuine unit root processes, nor genuine stationary processes. In other words, our case for nonlinear trend stationarity is based on circumstantial evidence.

8.2 Nonlinear co-trending test and estimation results

The components of the vector time series process z_t are now the CPIR and the FFR, for t = 1 (=1954.07) to 486 (=1994.12). The parameter α in Theorem 1 has been chosen equal to 1/2, for the reasons mentioned previously. Moreover, the tests are conducted on the basis of demeaned rather than detrended variables.

The ordered generalized eigenvalues of $\hat{M}_1$ w.r.t. $\hat{M}_2$ are $\hat{\lambda}_1 = 0.009134405$ and $\hat{\lambda}_2 = 0.04478581$, and the corresponding standardized generalized eigenvectors of $\hat{M}_1$ w.r.t. $\hat{M}_2$

5 The numerical results are not reported here. They can easily be replicated by using the author's software package *EasyReg* (see footnote 2), and the default test parameter options therein.

are

$$(30) \qquad \begin{matrix} 1 & -0.036827 & \leftarrow CPIR \\ -0.772864 & 1 & \leftarrow FFR \end{matrix}$$

Multiplying $\hat{\lambda}_r$ by $\sqrt{n} = \sqrt{486}$ now yields the test of the null hypothesis that there are *r* co-trending vectors against the alternative that there are less than *r* co-trending vectors. The test results, presented in Table 3, indicate that there is one co-trending vector.

Table 3: *Tests of the number r of co−trending vectors for* $z_t = (\text{CPIR}_t, \text{FFR}_t)^T$

r	*test statistic*	10% *crit. region*	5% *crit. region*	*conclusion*
1	0.20137	>0.35182	>0.46577	*accept*
2	0.98732	>0.53561	>0.67420	*reject*

In Figures 2 and 3 we display the components of the estimated functions $\hat{F}(x)$ and $\hat{F}'(x)$, respectively, standardized between -1 and 1 by dividing each component by its maximum absolute value. The common patterns in these components clearly corroborate the test result of presence of nonlinear co-trending.

<Insert Figures 2 and 3 about here>

It follows from Theorem 4 that the first column of (30) is a consistent estimate of this co-trending vector, and so is the normalized eigenvector of the matrix $\hat{M}_1$ corresponding to the smallest eigenvalue. The latter estimation result, which we shall adopt, reads:

$$(31) \qquad \textit{Nonlinear trend in } \text{CPIR} = 0.75457 \times \textit{Nonlinear trend in } \text{FFR}.$$

Note that the order in which we have written the nonlinear co-trending relation (31) should not be interpreted as a causal ordering, as each of the nonlinear trends in FFR and CPIR may be considered as the common nonlinear trend.

In order to determine the error of the estimate 0.75457, we have tested the hypothesis that the vector $H = (1,-a)^T$ is a co-trending vector, for *a* ranging from 0.3 to 1.2. It appears that the 95%

confidence interval of the parameter a is approximately (0.3, 1.2), and the 90% confidence interval is approximately (0.4, 1).

The case $a = 1$ is of particular interest, because it implies that the real interest rate has a constant expected value. The assumption of a constant ex-ante real interest rate (i.e., the difference between the nominal interest rate and the conditional expectation of the inflation rate), plays an important role in economic theory and finance. Our test results indicate that the hypothesis $a = 1$ is not rejected at the 5% significance level, and borderline (not) rejected at the 10% significance level. At first sight this result seems to contrast with Garcia and Perron (1996) and Phillips (1998), who find evidence of regime shifts in the real interest rate in the period 1961:1-1986:12. Garcia and Perron model the shifts in the mean of the real interest rate by a Hamilton (1989) type Markov switching model, with three states. Phillips tests for evidence of long memory in the same series. The interest rate in these two studies is the US 3 months Treasury bill rate. Therefore, we have redone the nonlinear co-trending analysis with the FFR replaced by this Treasury bill rate, and the US 4-6 months commercial paper rate, respectively. The results are about the same as for the federal funds rate, except that the estimate of a in both cases is now about 0.88, and the null hypothesis $a = 1$ is not rejected at the 10% significance level.

However, these results do not necessarily conflict with the findings of Garcia and Perron, because Hamilton's (1989) Markov switching model may generate data with a constant unconditional expectation and variance. Take for example the following AR(1) Markov switching model with two states, $S_t = 0$ and $S_t = 1$:

$$Y_t = \alpha S_t + \beta\left(Y_{t-1} - \alpha S_{t-1}\right) + \sigma e_t, \quad |\beta| < 1, \; e_t \sim i.i.d.\; N(0,1),$$
$$P(S_t = 1 | S_{t-1}=1) = p, \; P(S_t = 1 | S_{t-1} = 0) = q.$$

Then $E(Y_t) = \alpha E(S_t) = \alpha q/(1 - p + q)$. Therefore, the real interest rate can have a constant unconditional expectation, while being a AR(2) Markov switching process.

Phillips (1998) tests the fractional integration hypothesis $I(d)$ for the real interest rate, and finds that d is just outside the right-side of the stationary region (-0.5,0.5), but not significantly greater than 0.5 at the 5% significance level. Thus Phillips' results do not (clearly) conflict with ours.

9. THE PRICE PUZZLE RECONSIDERED

Our findings suggest that the positive correlation between the inflation rate and the interest rate is due to a common nonlinear deterministic time trend. However, does this common nonlinear trend explain the price puzzle? We recall that the price puzzle is the phenomenon that in certain VAR's the innovation response of inflation to a unit shock in interest is persistently positive. In Figure 4 we illustrate the price puzzle in a nonstructural VAR(36) with intercepts for the vector time series process $z_t = (\text{FFR}_t, \text{CPIR}_t)^T$ over a horizon of 60 months. The solid curve is the innovation response of the inflation rate to a unit shock in the federal funds rate, and the dotted curves are 1 and 2 times the standard error bands. These are asymptotic standard error bands, computed according to the approach of Baillie (1987).

<Insert Figure 4 about here>

We see that in this VAR the price puzzle is quite apparent: for the first 24 months the lower standard error band stays above the zero level, and although the innovation response curve dips below the zero level after 32 months, the negativity is not significant.

Next, we have re-estimated this VAR, but now including, next to the intercept, 20 Chebishev time polynomials $P_{j,n}(t)$, $j = 1,..,20$, in the two equations, in order to take (most of) the nonlinear trend out of the innovations. Chebishev time polynomials take the form

$$P_{0,n}(t) = 1, \quad P_{j,n}(t) = \sqrt{2}\cos[j\pi(t - 0.5)/n], \; j = 1,\dots,n-1.$$

See Hamming(1973) and Bierens (1997a). They are orthonormal: $(1/n)\sum_{t=1}^{n} P_{i,n}(t)P_{j,n}(t) = I(i \neq j)$ for $i,j = 0,...,n\text{-}1$, where $I()$ is the indicator function, and therefore any trend function $g(t)$ can be written as

$$g(t) = \sum_{j=0}^{n-1} \gamma_{j,n} P_{j,n}(t), \quad t = 1,..,n, \quad \textit{where} \;\; \gamma_{j,n} = (1/n)\sum_{t=1}^{n} g(t)P_{j,n}(t).$$

As can be seen from the graphs in Bierens (1997a), Chebishev polynomials are very flexible: If $g(t)$ is reasonably smooth, then for a relative (w.r.t. n) small number K, such as $K = 20$ in our case, the approximation $g_K(t) = \sum_{j=0}^{K} \gamma_{j,n} P_{j,n}(t)$ will likely be close to $g(t)$.

In both equations these Chebishev polynomials were jointly significant at the 5% level, corroborating the previous (circumstantial) evidence that both series are nonlinear trend stationary.

The effect of the inclusion of these Chebishev polynomials in the VAR on the innovation response of the CPIR is displayed in Figure 5.

<*Insert Figure 5 about here*>

We see from Figure 5 that the effect of this detrending procedure on the innovation response of the CPIR is quite astonishing. The innovation response curve starts off significantly positive only for a few months, and then wiggles around the zero level. Comparing Figures 4 and 5 we also see that the magnitude of the response has been substantially reduced. Therefore, the conclusion seems justified that the price puzzle is, to a large extent, due to a common nonlinear trend in the FFR and the CPIR. This conclusion, of course, does not mean that the price puzzle is solved! Another (price) puzzle has now emerged:

10. WHY IS THERE A COMMON NONLINEAR TREND IN THE CPI INFLATION RATE AND THE FEDERAL FUNDS RATE?

As has already been alluded to in Section 1, the shapes of the standardized CPI inflation rate (CPIR) and the standardized inflation rate of the PPI of fuel and related products[6] (PPIFIR), plotted in Figure 6, have a remarkable resemblance. The sharp rise in the CPIR in the early seventies and around 1980 coincide with the rise of the PPIFIR. In particular, the dates of the top of the two main peaks in the PPIFIR are two to three months earlier than those of the CPIR. These peaks are clearly the results of the two oil price shocks induced by the OPEC cartel.

<*Insert Figure 6 about here*>

In order to verify whether the PPIFIR is the source of the common nonlinear trend in the CPIR and the FFR, we have conducted our co-trending analysis in the same way as before on the vector time series process $z_t = (FFR_t, CPIR_t, PPIFIR_t)^T$, after having tested for a possible unit root in the PPIFIR (using all available data, from 1947.01 to 1998.12). As for the latter, the results were again mixed. The Phillips-Perron test rejects the unit root hypothesis for PPIFIR at the 5% significance level, and so do the HOAC(1,1) and HOAC(2,2) tests of Bierens (1993). The Bierens-Guo (1993) stationarity tests BG(1) and BG(2) reject the stationarity hypothesis at the 5%

[6] The original source is URL http://www.stls.frb.org/fred/data/ppi/ppieng of the Federal Reserve Bank of St. Louis.

significance level. However, the tests BG(3) and BG(4) do not reject the stationarity hypothesis at the 10% significance level, and the same applies to the KPSS test. Therefore, also this time series is neither a genuine unit root process, nor a genuine stationary process.

Table 4: Tests of the number r of cotrending vectors for $z_t = (\text{CPIR}_t, \text{FFR}_t, \text{PPIFIR}_t)^T$				
r	*test statistic*	10% *crit. region*	5% *crit. region*	*conclusion*
1	0.04529	>0.35183	>0.46577	*accept*
2	0.21503	>0.53561	>0.67420	*accept*
3	1.29460	>0.70366	>0.86038	*reject*

As expected, the co-trending test results in Table 4 indicate that there are two co-trending vectors: r = 2, with estimation results (based on the eigenvalues of $\hat{M}_1$):

$$\begin{aligned} &\textit{Nonlinear trend in } \text{FFR} = 0.459726 \times \textit{Nonlinear trend in } \text{PPIFIR} \\ &\textit{Nonlinear trend in } \text{CPIR} = 0.430217 \times \textit{Nonlinear trend in } \text{PPIFIR} \end{aligned} \tag{32}$$

We have tested again whether the real FFR is stationary, by testing whether the 3×1 matrix $H = (-1,1,0)^T$ spans a subspace of the space of co-trending vectors. Note that this is just the test of the null hypothesis that the two coefficients in (32) are equal. The test result is slightly different from before: the null hypothesis involved is still not rejected at the 5% significance level, but it is now rejected at the 10% significance level. This result is therefore more in tune with the results of Garcia and Perron (1996) and Phillips (1998) than before.

In view of these results, the conclusion seems justified that the non-linear trend in the CPIR is to a large extent due to the nonlinear trend in the PPIFIR, which in its turn is to a large extent due to the oil price shocks and their aftermath, induced by OPEC.

As to the nonlinear trend in the FFR, it is likely that the FOMC of the Federal Reserve Board has anticipated the inflationary effect of the actions of the OPEC, and responded by preemptive raises of the FFR. Moreover, it should be (and probably has been) a matter of concern to the FOMC if the real interest rate runs out of hand, because too high a real interest rate will have serious negative effects on the real economy, and too low a real interest rate will boost spending on credit and therefore inflation. The FOMC will therefore likely keep the real interest rate within certain bounds, which may (partly) explain the common nonlinear trend in the FFR and the CPIR.

ACKNOWLEDGEMENTS

Previous versions of this paper have been presented at York University, Texas Econometrics Camp 1996, Pennsylvania State University, Econometric Society European Meeting 1996, University of Pennsylvania, New York University, University of Maryland, Tilburg University, and the University of Southern California. The useful comments of Dave Cushman and two referees are gratefully acknowledged.

REFERENCES

Anderson, S.A., H.K. Brons, and S.T. Jensen (1983), "Distribution of Eigenvalues in Multivariate Statistical Analysis", *Annals of Statistics*, **11**, 392-415.

Baillie, R.T. (1987), "Inference in Dynamic Models Containing 'Surprise' Variables", *Journal of Econometrics*, **35**, 101-117.

Balke, N.S. and K.M. Emery (1994), "The Federal Funds Rate as an Indicator of Monetary Policy: Evidence from the 1980's", *Federal Reserve Bank of Dallas Economic Review*, 1-15

Balke, N.S. and K.M. Emery (1995), "Understanding the Price Puzzle", Working paper, Department of Economics, Southern Methodist University

Bernanke, B.S. and A.S. Blinder (1992), "The Federal Funds Rate and the Channels of Monetary Transmission", *American Economic Review*, **82**, 901-921.

Bierens, H.J. (1993), "Higher-order Sample Autocorrelations and the Unit Root Hypothesis", *Journal of Econometrics*, **57**, 137-160.

Bierens, H.J. (1994), *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models* (Cambridge, U.K.: Cambridge University Press).

Bierens, H.J. (1997a), "Testing the Unit Root with Drift Hypothesis Against Nonlinear Trend Stationarity, with an Application to the U.S. Price Level and Interest Rate", *Journal of Econometrics*, **81**, 29-64.

Bierens, H.J. (1997b), "Nonparametric Cointegration Analysis", *Journal of Econometrics*, **77**, 379-404.

Bierens, H.J. and S. Guo (1993), "Testing Stationarity and Trend Stationarity Against the Unit Root Hypothesis", *Econometric Reviews*, **12**, 1-32.

Bierens, H.J. and W. Ploberger (1997), "Asymptotic Theory of Integrated Conditional Moment Tests", *Econometrica*, **65**, 1129-1151.

Billingsley, P. (1968), *Convergence of Probability Measures* (New York: John Wiley).

Christiano, L.J. and M. Eichenbaum (1992), "Identification and the Liquidity Effect of Monetary Policy Shocks", in A. Cuikerman, Z. Hercowitz and L. Leiderman (Eds), *Political Economy, Growth, and Business Cycles*. (Cambridge, Massachusetts: M.I.T. Press)

Christiano, L.J., M. Eigenbaum, and C. Evans (1994), "The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds", Federal Reserve Bank of Chicago Working Paper Series 94-2

Christiano, L.J., M. Eichenbaum, and C. Evans (1995), "Identification and the Effect of Monetary Policy Shocks", in: M. Blejer, Z. Eckstein, Z. Hercowitz, and L. Leiderman (Eds), *Factors in Economic Stabilization and Growth*. (Cambridge, U.K.: Cambridge University Press)

Cushman, D.O., and T. Zha (1997), "Identifying Monetary Policy in a Small Open Economy Under Flexible Exchange Rates", *Journal of Monetary Economics*, **39**, 433-448.

Dunford, N. and J.T. Schwartz (1963), *Linear Operators, Part II: Spectral Theory* (New York: Wiley Interscience).

Eichenbaum, M. (1992), " Comment on 'Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy' by Christopher Sims", *European Economic Review*, **36**, 1001-1011.

Engle, R.F. (1987), "On the Theory of Cointegrated Economic Time Series", Invited paper presented at the Econometric Society European Meeting 1987, Copenhagen.

Engle, R.F. and S. Kozicki (1993), "Testing for Common Features", *Journal of Business and Economic Statistics*, **11**, 369-386.

Garcia, R. and P. Perron (1996), "An Analysis of the Real Interest Rate Under Regime Shifts", *Review of Economics and Statistics*, 111-125

Hall, P. and C.C. Heyde (1980), *Martingale Limit Theory and Its Applications* (San Diego: Academic Press).

Hamilton, J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica* 57, 357-384.

Hamming, R.W. (1973), *Numerical Methods for Scientists and Engineers* (New York: Dover Publications).

Johansen, S. (1988), "Statistical Analysis of Cointegrated Vectors", *Journal of Economic Dynamics and Control*, **12**, 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrated Vectors in Gaussian Vector Autoregressive Models", *Econometrica,* **59**, 1551-1580.

Johansen, S. (1994), "The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables", *Econometric Reviews*, **13**, 205-230.

Johansen, S. and K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration, with Applications to the Demand for Money", *Oxford Bulletin of Economics and Statistics,* **52**, 169-210.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992), "Testing the Null of Stationarity Against the Alternative of a Unit Root", *Journal of Econometrics*, **54**, 159-178.

Lee, D. and P. Schmidt (1996), "On the Power of the KPSS Test of Stationarity Against Fractionally-Integrated Alternatives", *Journal of Econometrics*, **73**, 285-302.

Nelson, C.R. and C.I. Plosser (1982), "Trends and Random Walks in Macro-economic Time Series", *Journal of Monetary Economics,* **10**, 139-162.

Newey, W.K. and K.D. West (1987), "A Simple Positive Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, **55**, 703-708

Perron, P. (1988), "Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach", *Journal of Economic Dynamics and Control,* **12**, 297-332.

Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", *Econometrica*, **57**, 1361-1402.

Perron, P. (1990), "Testing the Unit Root in a Time Series with a Changing Mean", *Journal of Business and Economic Statistics*, **8**, 153-162.

Phillips, P.C.B. (1998), "Econometric Analysis of Fisher's Equation", Working Paper, Department of Economics, Yale University.

Phillips, P.C.B. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, **75**, 335-346.

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

Schotman, P.C. and H.K. van Dijk (1991), "On Bayesian Routes to Unit Roots", *Journal of Applied Econometrics*, **6**, 387-401.

Sims, C.A. (1992), "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy", *European Economic Review*, **36**, 975-1000.

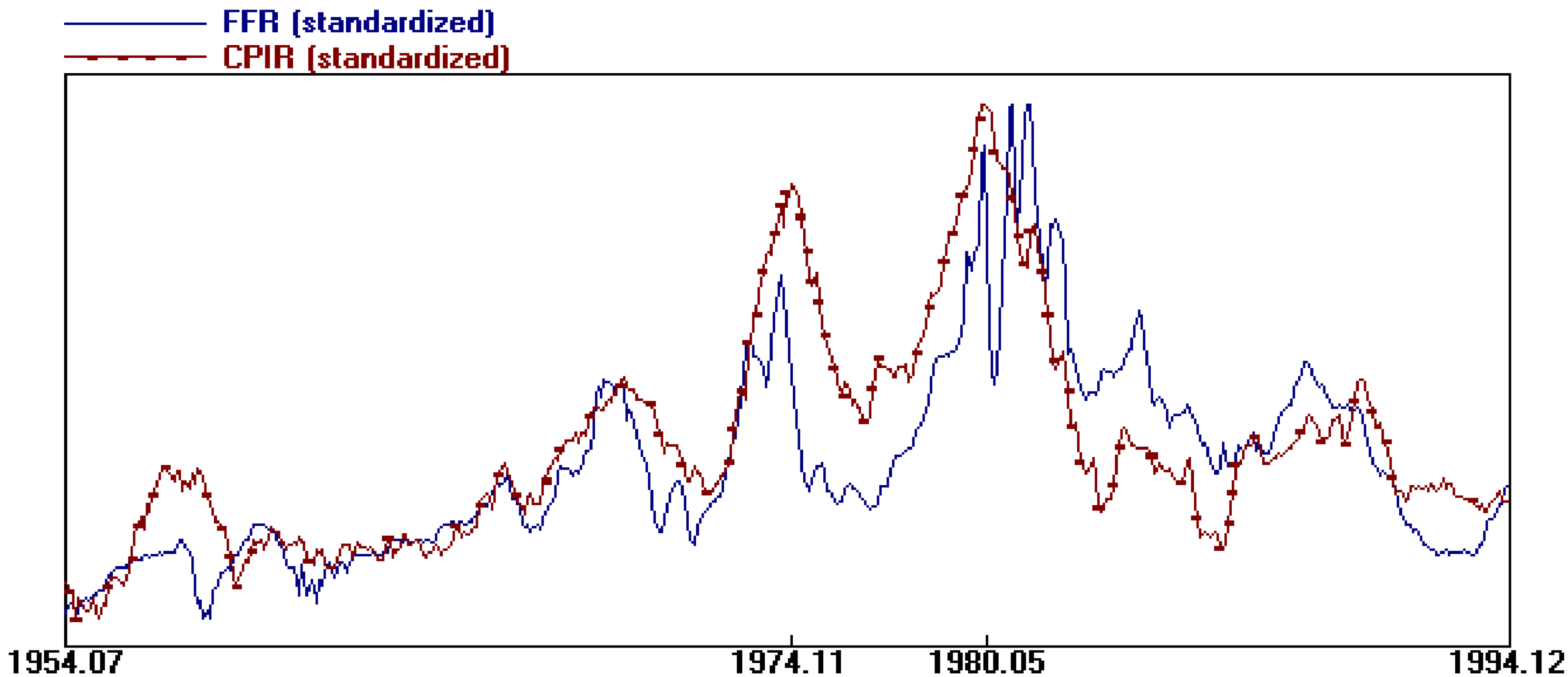

Figure 1: The data

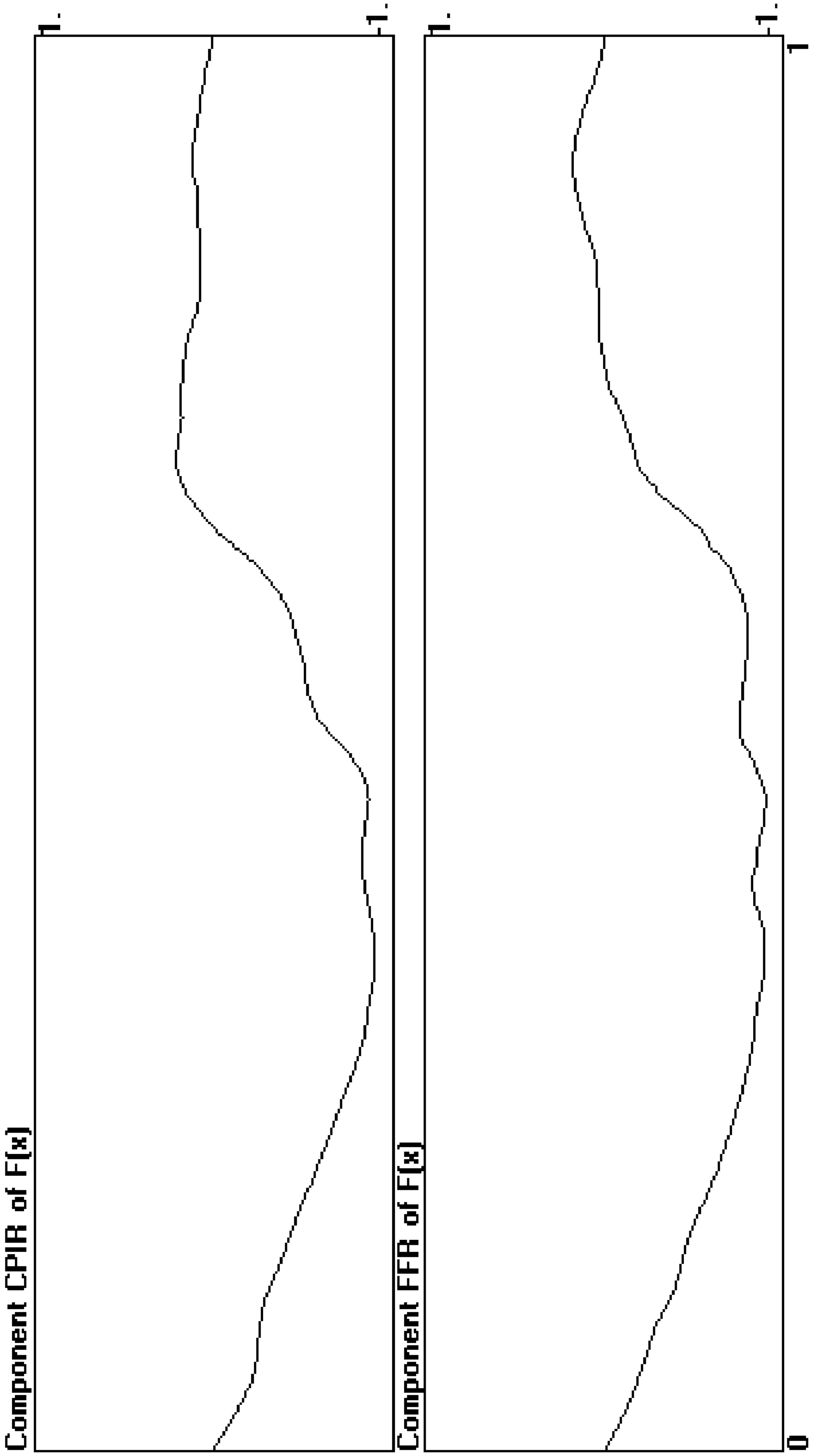

Figure 2: Rescaled components of F(x)

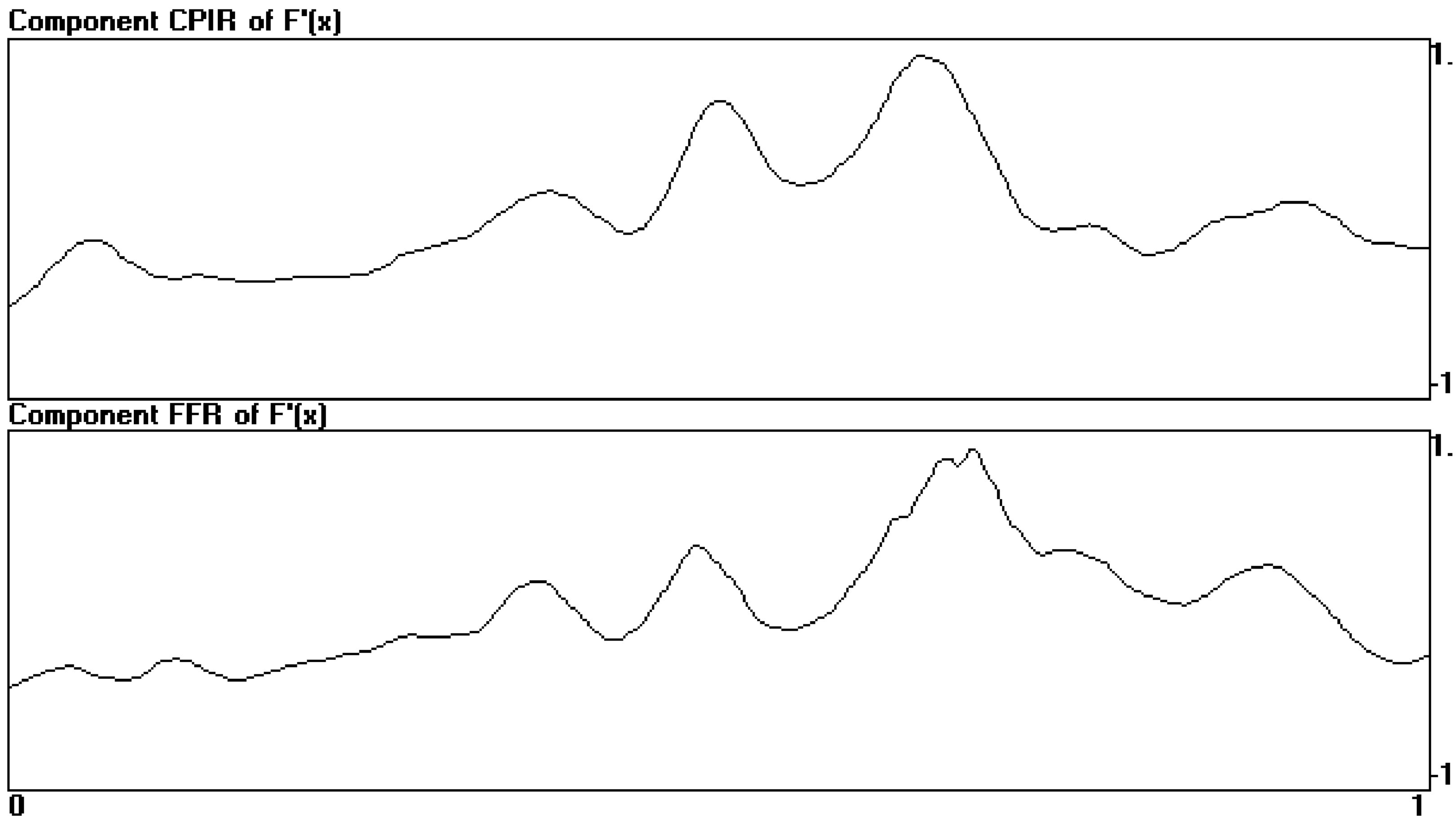

Figure 3: Rescaled components of F'(x)

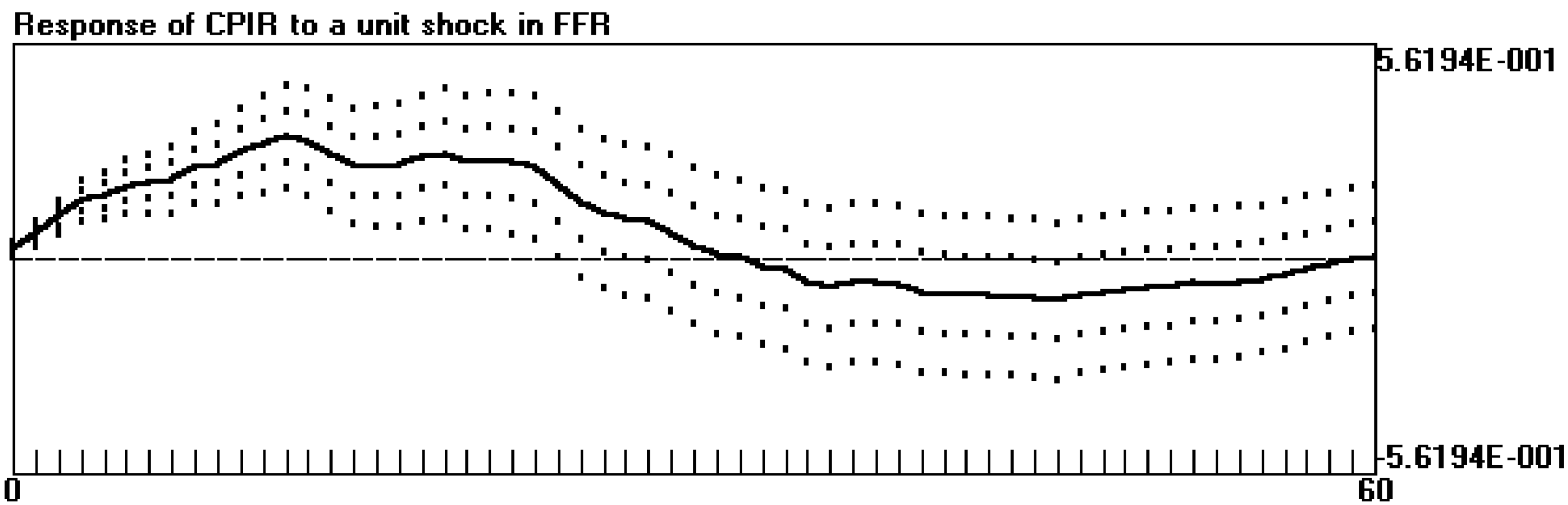

Figure 4: The price puzzle

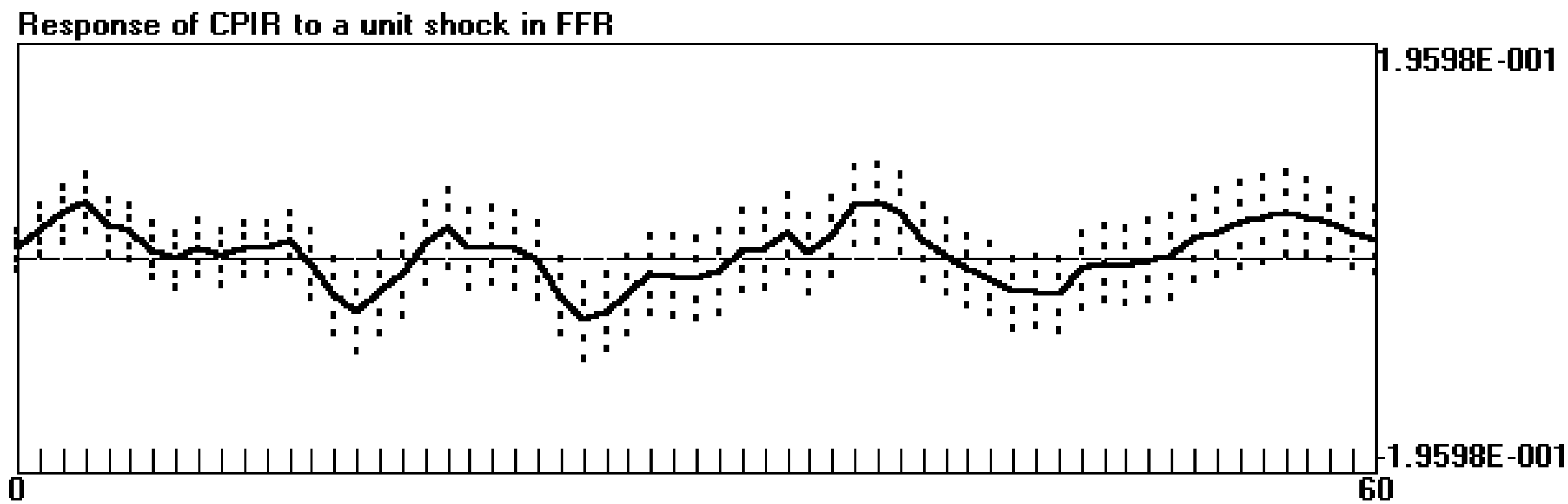

Figure 5: The price puzzle after taking out the nonlinear time trend

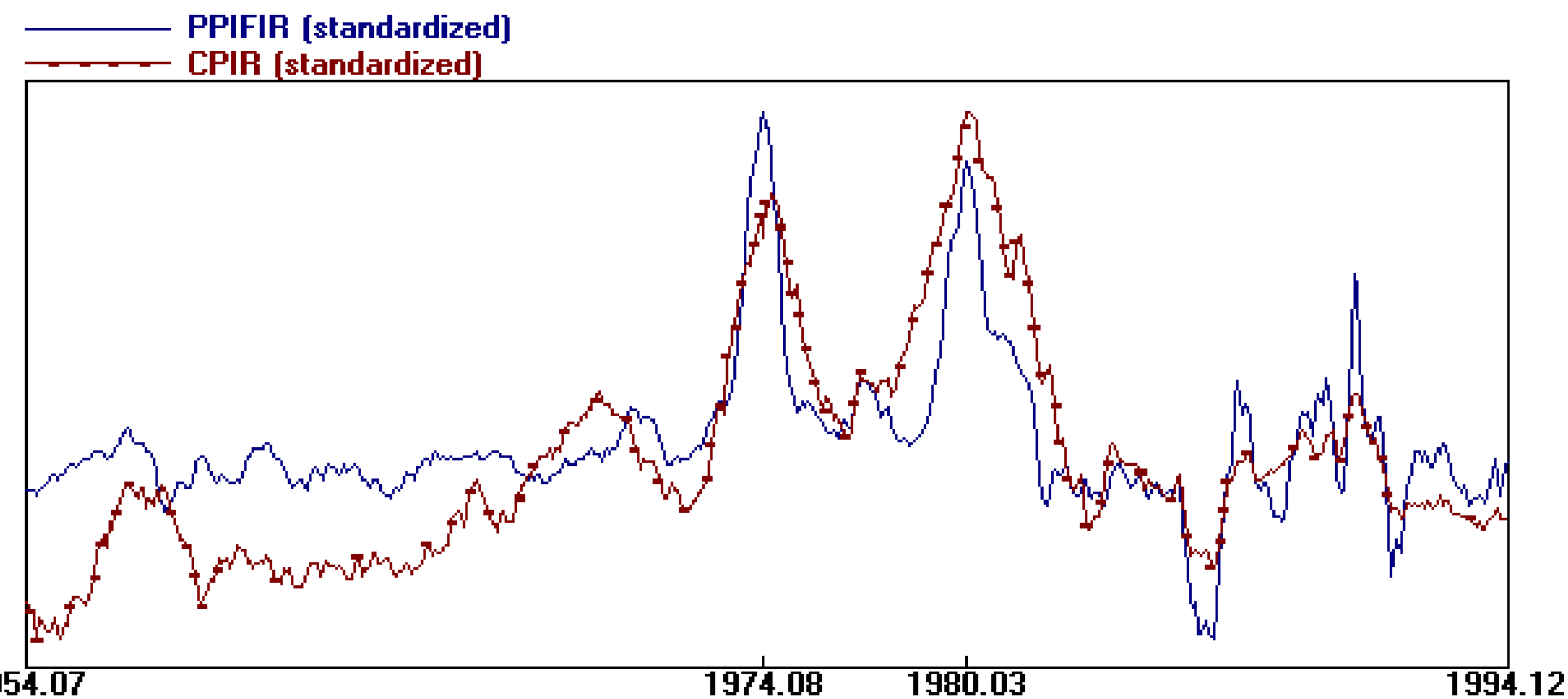

Figure 6: PPIFIR compared with CPIR