

NONPARAMETRIC FACTOR ANALYSIS OF TIME SERIES

Juan M. Rodriguez-Poo*, Universidad de Cantabria
and Oliver Linton, Yale University

March 19, 1998

Abstract

We introduce a nonparametric smoothing procedure for nonparametric factor analysis of multivariate time series. The asymptotic properties of the proposed procedures are derived. We present an application based on the residuals from the Fair macromodel.

Keywords. Factor Analysis; Kernel estimation; Nonparametric; Time Series.

*Address correspondence to: Juan M. Rodriguez-Poo. Departamento de Economia. Universidad de Cantabria. Avda. de los Castros s/n. 39005 Santander. Spain. e-mail: rodrigjm@ccaix3.unican.es .

1 Introduction

The residuals from many time series models have provided evidence of time varying second moments. In the univariate context, the ARCH family of models are widely used to reflect this phenomenon following the seminal work of Engle (1982). These nonlinear models appear to work quite well in terms of explaining volatility clustering and heavy tailed error distributions. Nevertheless, they are completely arbitrary as is evidenced by the rapidly increasing variations on this theme, including: GARCH, EGARCH, and STARARCH. In the multivariate context, these models are difficult to apply due to the parameter explosion: unrestricted multivariate ARCH contains $O(d^4)$ parameters, where d is the dimensionality. Several restricted versions of these models have found favour including factor GARCH. See Bollerslev, Engle and Nelson (1994) for a nice review.

We eschew the overly strong restrictions imposed by these parametric models in favour of a nonparametric approach to modelling the second moment structure. We merely assume that the conditional covariance matrix, $\Sigma(t)$, is a smooth function of time. We introduce a method for analyzing and exploiting the structure in the conditional covariance matrix of multivariate time series. We apply standard multivariate techniques at each time point, as in Anderson (1984), by exploiting a hypothesized smoothness of the second moment matrix over time. This covariance matrix can be estimated by a multitude of nonparametric smoothing techniques, see Härdle and Linton (1994), without generally requiring the nonlinear optimization on which ARCH modellers must rely. We also estimate the eigenvalues of $\Sigma(t)$, and use these to provide several tests that the covariance matrix is of reduced rank somewhere in the sample. Our procedures can be used more generally to quantify the changing second moment structure in the data and to improve efficiency of semiparametric procedures as in Robinson (1987). See Donald (1995) for some applications of nonparametric factor analysis in a cross-sectional setting.

We apply our procedures to the residuals of the well-known Fair-model, which is a 30 equation macro-model of the US economy, and find substantial evidence of time-varying second moments as well as changing structure in the cross-relationship between the errors of that system.

The paper is organized as follows. In section 2 we present the statistical model and some preliminary results. In section 3, we show the main results concerning the asymptotic distribution of the

covariance matrix estimator and the eigenvalues. In section 4 we develop some testing procedures and finally in section 5 we present an application. The proof of the main results are relegated to the Appendix.

2 Preliminaries

We work inside the fairly general framework of a nonlinear simultaneous equations model

$$y_t = g(y_t^{**}, x_t; \theta) + u_t, \quad t = 1, 2, \dots, T, \quad (2.1)$$

where $y_t = (y_t^*, y_t^{**})^T$ is a $d \times 1$ vector of dependent variables, x_t is a k vector of explanatory variables and the function $g(\cdot)$ depends only on a p vector θ of unknown parameters. We shall assume that the model is correctly specified in so far as the errors u_t are martingale difference sequences with respect to the natural filtration $\mathcal{F}_t = \sigma(x_t, u_{t-1}, \dots)$. We will also maintain throughout that the covariance matrix of u_t , $\Sigma(t)$, is a smooth function of t . Suppose that we have a vector of consistent estimates, $\hat{\theta}_T$, of θ , available from instrumental variable techniques, perhaps. Define the vector of residuals

$$\hat{u}_t = y_t - g(y_t^{**}, x_t; \hat{\theta}_T), \quad t = 1, 2, \dots, T.$$

If the parametric model (2.1) is correctly specified, these residuals approximate well the unobservable errors. This framework is broadly consistent with the model estimated in Fair (1994).

Our main interest here is in estimating the unknown function $\Sigma(t) = (\sigma_{lk}(t))_{l,k}$, and functions thereof, given the residual vectors $\{\hat{u}_t\}_{t=1}^T$. We estimate $\sigma_{lk}(t)$ by

$$\hat{\sigma}_{lk}(t) = \frac{1}{T} \sum_{s=1}^T K_h(t-s) \hat{u}_{ls} \hat{u}_{ks} - \hat{m}_l(t) \hat{m}_k(t), \quad l, k = 1, \dots, d, \quad (2.2)$$

where $\hat{m}_k(t) = T^{-1} \sum_{s=1}^T K_h(t-s) \hat{u}_{ks}$, in which $K_h(\cdot) = h^{-1} K(\cdot/Th)$ denotes the rescaled kernel function with $\int K(u) du = 1$, and the bandwidth sequence $h(T) \downarrow 0$. Our method also works for unequally spaced observations at time points $\{t_s\}_{s=1}^T$, in which case we would take the following weights $K\left(\frac{t-t_s}{Th}\right) / \sum_{r=1}^T K\left(\frac{t-t_r}{Th}\right)$, $s = 1, \dots, T$. Now let $\hat{\Sigma}(t) = (\hat{\sigma}_{kl}(t))_{k,l}$ and let $\hat{\lambda}_1(t), \dots, \hat{\lambda}_d(t)$ be the eigenvalues of the real symmetric matrix $\hat{\Sigma}(t)$, sorted in decreasing order of magnitude. We can estimate any smooth function of $\Sigma(t)$ by the analogous function of $\hat{\Sigma}(t)$; for example, we estimate $\det(\Sigma(t))$ by $\det(\hat{\Sigma}(t))$ and the trace ratio

$$\psi_k(t) = \frac{\lambda_1(t) + \cdots + \lambda_k(t)}{\lambda_1(t) + \cdots + \lambda_d(t)} \quad \text{by} \quad \hat{\psi}_p(t) = \frac{\hat{\lambda}_1(t) + \cdots + \hat{\lambda}_p(t)}{\hat{\lambda}_1(t) + \cdots + \hat{\lambda}_d(t)}.$$

Both of these functions are of particular interest when analyzing the multivariate structure present in the data. Specifically, whether some of the eigenvalues are zero for some subperiods.

REMARKS.

1. The estimates of $\Sigma(t)$ are linear in the squared residuals and their lags with declining weights. This is as in the ARCH case in which the estimate for the conditional variance is itself a linear combination of the squared residuals.
2. Estimating the means by $\widehat{m}_k(t)$ is not strictly necessary when the parametric model for the mean is correctly specified.

3 Main Results

Let $\|A\| = \text{tr}^{1/2}(A'A)$ be the Euclidean norm of any matrix A . In order to show the main results of our paper we need the following assumptions.

(H.1) The data generating process $\{y_t, x_t\}_{t=1}^T$ is a weakly stationary array of random variables and $E|y_t|^r < \infty$. Moreover, it is strongly mixing with coefficients that satisfy $\sum_{k=1}^{\infty} \alpha(k)^{1-\frac{2}{r}} < \infty$, for $r > 2$.

(H.2)

$$E[y_t | y_{t-1}, y_{t-2}, \dots, x_t] = g(y_t^{**}, x_t; \theta_0) \quad \text{a.s.}$$

(H.3) $\text{var}(y_t) = \Sigma(t)$. The elements of $\Sigma(t)$ have bounded first derivatives.

(H.4) Let

$$\beta_{ll'k}(t) = E[y_{lt}y_{l't}y_{kt}] \quad \text{and} \quad \gamma_{ll'kk'}(t) = E[y_{lt}y_{l't}y_{kt}y_{k't}].$$

The functions $\beta_{ll'k}(t)$ and $\gamma_{ll'kk'}(t)$ have bounded first derivatives, for $l, l', k, k' = 1, \dots, d$

(H.5) $\text{cov}(|y_t|^{r_1}, |y_t|^{r_2}) < \infty$ for $r_1, r_2 = 0, 1, 2, 3$.

(H.6) $T^{1/2}(\hat{\theta}_T - \theta)$ is stochastically bounded, i.e., for all $\epsilon > 0$ there exists $M < \infty$ such that $\Pr \left[\left\| T^{1/2}(\hat{\theta}_T - \theta) \right\| > M \right] < \epsilon$.

(H.7) The regression function $g(\bullet; \theta)$ is differentiable in θ on a neighborhood Θ_1 of θ_0 a.s. Moreover,

$$\sup_{T \geq 1} \frac{1}{T} \sum_{s=1}^T E \sup_{\theta \in \Theta_1} \left\| \left[\frac{\partial}{\partial \theta} g(y_s^{**}, x_s; \theta^*) \otimes \frac{\partial}{\partial \theta} g(y_s^{**}, x_s; \theta^*)^T \right] \right\| < \infty,$$

and

$$\sup_{T \geq 1} \frac{1}{T} \sum_{s=1}^T E \sup_{\theta \in \Theta_1} \left\| \left[\frac{\partial}{\partial \theta} g(y_s^{**}, x_s; \theta^*) \otimes I \right] \right\| < \infty.$$

(H.8) The kernels $K(u)$, $|uK(u)|$ and $\sup_u |K(u)|$ are bounded.

(H.9) The first moment of the kernel is zero, $\int uK(u)du = 0$.

(H.10) $\int_0^1 K^2(u)du < \infty$, and the second kernel moment is bounded, $\int u^2 K(u)du < \infty$.

With these definitions and assumptions, we establish the following results in the appendix:

PROPOSITION 1. *Under assumptions (H.1) to (H.10) and if the bandwidth h tends to zero, such that $Th^5 \rightarrow 0$ and $Th^3 \rightarrow \infty$, then as T tends to infinity,*

$$(Th)^{1/2} \left\{ \text{vech} \left[\hat{\Sigma}(t) \right] - \text{vech} \left[\Sigma(t) \right] \right\} \Longrightarrow N \{0, V(t)\},$$

where $V(t) = \{v_{lk, l'k'}(t)\}$ with $q = \frac{d(d+1)}{2}$ columns and rows, where

$$v_{lk, l'k'}(t) = \int K^2(u)du \times [\text{cov}\{\aleph_{lk}(t), \aleph_{l'k'}(t)\}]$$

and $\aleph_{lk}(t) = u_{lt}u_{kt}$. Furthermore, $\hat{V}(t) = \{\hat{v}_{lk, l'k'}(t)\} \rightarrow_p V(t)$, where

$$\hat{v}_{lk,l'k'}(t) = \int K^2(u)du \times \frac{1}{T} \sum_{s=1}^T K_h(t-s) \{\hat{u}_{ls}\hat{u}_{ks} - \hat{\sigma}_{lk}(t)\} \{\hat{u}_{l's}\hat{u}_{k's} - \hat{\sigma}_{l'k'}(t)\}.$$

In order to formulate the limit result for the estimated eigenvalue vector we need some more notation. Let $D_t(\lambda) = \det \{\Sigma(t) - \lambda I\}$ denote the characteristic polynomial of the covariance matrix $\Sigma(t)$. Let $B_{lk}(t)$ be the cofactor of the (l, k) -th element of $\Sigma(t) - \lambda(t)I$, and let

$$B(\lambda, t) = 2 \{B_{lk}(\lambda, t)\}_{lk} - \text{diag} \{B_{11}(\lambda, t), B_{22}(\lambda, t), \dots, B_{dd}(\lambda, t)\}.$$

The matrix $\{B_{lk}(\lambda, t)\}_{lk}$ consists of all cofactors of $\Sigma(t) - \lambda I$. With this matrix we also define a $q \times d$ matrix

$$M(t) = \left[\frac{\text{vech}(B(\lambda_1, t))}{D'_t(\lambda_1(t))}, \frac{\text{vech}(B(\lambda_2, t))}{D'_t(\lambda_2(t))}, \dots, \frac{\text{vech}(B(\lambda_d, t))}{D'_t(\lambda_d(t))} \right],$$

and finally $\Omega(t) = M(t)^T V(t) M(t)$. Let the corresponding quantities derived from $\hat{\Sigma}(t)$ be indicated by hat superscripts, e.g., $\hat{D}_t(\lambda) = \det \{\hat{\Sigma}(t) - \lambda I\}$.

With this notation, we can now formulate the asymptotic distribution of the vector of eigenvalues $\hat{\lambda}(t) = (\hat{\lambda}_1(t), \hat{\lambda}_2(t), \dots, \hat{\lambda}_d(t))^T$ of the estimated covariance matrix $\hat{\Sigma}(t)$.

PROPOSITION 2. *Under assumptions (H.1) to (H.10) and if the bandwidth h tends to zero, such that $Th^5 \rightarrow 0$ and $Th^3 \rightarrow \infty$, then*

$$(Th)^{1/2} \{\hat{\lambda}(t) - \lambda(t)\} \implies N \{0, \Omega(t)\},$$

as T tends to infinity. Furthermore, $\hat{\Omega}(t) \rightarrow_p \Omega(t)$, where $\hat{\Omega}(t) = \hat{M}(t)^T \hat{V}(t) \hat{M}(t)$.

4 Hypothesis Testing

There are a number of hypotheses of interest here. First, one might be interested in testing for independence or conditional independence (correlation) between two subsets of the variables. Second, one may like to summarize the covariance matrix by a smaller number of variables through

principal components analysis, canonical correlations, or factor analysis, see Anderson (1984). This dimensionality reduction aids interpretation. For this purpose we may want to test whether some of the eigenvalues are zero. Finally, we may want to examine the time variation of statistics that measure discrepancy from the above null hypotheses in order to give information about structural stability of the underlying model.

4.1 Independence Tests

Let

$$\Sigma(t) = \begin{pmatrix} \Sigma_{11}(t) & \Sigma_{12}(t) \\ \Sigma_{21}(t) & \Sigma_{22}(t) \end{pmatrix} \quad ; \quad \Sigma^{-1}(t) = \begin{pmatrix} \Sigma^{11}(t) & \Sigma^{12}(t) \\ \Sigma^{21}(t) & \Sigma^{22}(t) \end{pmatrix},$$

where $\Sigma_{11}(t)$ and $\Sigma^{11}(t)$ are conformable $k \times k$ submatrices, for some $k \leq d$. We wish to test the following hypotheses

$$H_0^1 : \Sigma_{12}(t) = 0 \quad ; \quad H_0^2 : \Sigma^{12}(t) = 0.$$

These hypotheses are of interest because their veracity would indicate that all dependencies across sectors have been accounted for by the primary model.

The following local coefficients of alienation can be used to measure departures from the above hypotheses

$$\hat{\tau}_1(t) = \frac{\det(\hat{\Sigma}(t))}{\det(\hat{\Sigma}_{11}(t)) \det(\hat{\Sigma}_{22}(t))} - 1 \quad ; \quad \hat{\tau}_2(t) = \frac{\det(\hat{\Sigma}^{-1}(t))}{\det(\hat{\Sigma}^{11}(t)) \det(\hat{\Sigma}^{22}(t))} - 1.$$

Their asymptotic distributions follow from Propositions 1 and 2, i.e.,

$$(Th)^{1/2} \{ \hat{\tau}_j(t) - \tau_j(t) \} \implies N \{ 0, G(t)^T V(t) G_j(t) \}, \quad (1)$$

where $G_j(t) = \partial \tau_j(t) / \partial \text{vech}(\Sigma(t))$.

4.2 Reduced Rank Tests

We are interested in the following null hypotheses

$$H_0^3 : \det(\Sigma(t)) = 0 \quad ; \quad H_0^4 : \psi_p(t) = \frac{\lambda_1(t) + \cdots + \lambda_p(t)}{\lambda_1(t) + \cdots + \lambda_d(t)} = 0.$$

In the first case, we are testing for reduced rank in the covariance matrix. The second test also establishes by how much the rank is reduced. If such structure is present in the data, it may be advisable to modify ones estimation procedures for θ , or even to respecify the underlying model. Informally, these quantities provide useful diagnostic evidence about the validity of a model.

Denote by $C_{lk}(t)$ the cofactor of the (l, k) 'th element of $\Sigma(t)$, and let

$$C(t) = 2 \{C_{lk}(t)\}_{lk} - \text{diag} \{C_{11}(t), C_{22}(t), \dots, C_{dd}(t)\},$$

where $\{C_{lk}(t)\}_{lk}$ is the matrix of all cofactors of $\Sigma(t)$. Then, $\det(\Sigma)$ is a smooth function in t with

$$\frac{\partial \det(\Sigma)}{\partial \text{vech}(\Sigma)} = \text{vech}(C).$$

Therefore, by Theorem A, p. 122 of Serfling (1980) we have

$$(Th)^{1/2} \left\{ \det(\widehat{\Sigma}(t)) - \det(\Sigma(t)) \right\} \implies N \left\{ 0, \text{vech}[C(t)]^T V(t) \text{vech}[C(t)] \right\}. \quad (2)$$

Finally, define the vector of d components

$$P_i(t) = \begin{cases} \frac{1 - \psi_i(t)}{\text{tr}(\Sigma)} & i = 1, \dots, k \\ \frac{-\psi_i(t)}{\text{tr}(\Sigma)} & i = k + 1, \dots, d, \end{cases}$$

and then

$$(Th)^{1/2} \left\{ \widehat{\psi}_i(t) - \psi_i(t) \right\} \implies N \left\{ 0, P_i(t)^T \Omega(t) P_i(t) \right\}. \quad (3)$$

In cases (1-3), consistent standard errors can be constructed in the manner of Theorem 1 and 2.

5 Application to the Fair Model

The Fair model is a 28 equation model of the US economy implemented with quarterly data from 1975 to the present day. The model is composed of five sectors: households, firms, financial, import and government sectors. There are several issues of interest. First, we want to quantify the time-varying second moments. We estimate the conditional variances of all equation errors jointly with their pointwise confidence bands. In Figure 1(a,b,c,d,e), we show how the conditional variances of the error terms in equations 1 (household sector), 10 (firm sector), 22 (financial sector), 27 (import equation) and 28 (government sector) vary over time.

Figure 1(a,b,c,d,e) is about here

In all cases it is clear that second order moments exhibit a time varying path. However, the conditional variances do not give to much information about the structure of the covariance matrix. We would like to know for example, if the 28 equations of the Fair model are linearly independent or, on the contrary, if there exists colinearity among them. One possible way to analyze the structure of this covariance matrix is through the analysis of the eigenvalues. We can test for the null hypothesis that some eigenvalues are zero. Using the estimator developed in Proposition 2, we have computed the eigenvalues of the covariance matrix of the Fair model as a function of a time index. The results of this estimation procedure are shown in Figure 2

Figure 2 is about here

It is hard to obtain some meaningful information from the time-plot of the eigenvalues, however, there are several possibilities to summarize this information and to make it more useful. Following the results obtained in Proposition 3, we have computed both the determinant of the covariance matrix and its pointwise confidence bands. In Figure 3 we show the determinant of the covariance matrix estimated at several time values.

Figure 3 is about here

Nowhere in the sample it is possible to reject the null hypothesis that the determinant of the covariance matrix is equal to zero at a 95 percent of significance level. Therefore, there exists some collinearity in the structure of the covariance matrix. In order to know more about the degree of this reduction of dimensionality, we have computed the trace ratio proposed in Proposition 4 for $k = 2, 3, 4, 5, 6$ and 7 jointly with their pointwise confidence bands. The values of the trace-ratio statistics jointly with the confidence bands are shown in Figure 4(a,b,c,d,e,f). The smallest three eigenvalues of the covariance matrix are not significantly different from zero. Only up to the sixth eigenvalue the trace-ratio is significantly different from one everywhere in the sample.

Figure 4(a,b,c,d,e,f) is about here

Finally, it is also of interest to test for the independence of the five sectors included in the Fair model pairwise. In Figure 5, we show the statistic $\tau_1(t)$ computed for the five sectors and several time values.

Figure 5 is about here

Only the import block of equations shows independence with respect the other blocks and therefore it would be possible to estimate individually this block with respect to the others without efficiency loss.

6 Appendix

The following lemmas are used in the proof of Proposition 1.

LEMMA 1. *Let*

$$\tilde{\Sigma}(t) = T^{-1} \sum_{s=1}^T K_h(t-s) u_s u_s^T - \tilde{m}(t) \tilde{m}(t)^T$$

where $\tilde{m}(t) = T^{-1} \sum_{s=1}^T K_h(t-s) u_s$ and $u_s = y_s - g(y_s^{**}, x_s, \theta_0)$ for $t = 1, \dots, T$. Then, under assumptions (H.1) to (H.4), if $Th^5 \rightarrow 0$ and $Th^3 \rightarrow \infty$ as T tends to infinity,

$$(Th)^{1/2} \left\{ \text{vech} [\tilde{\Sigma}(t)] - \text{vech} [\Sigma(t)] \right\} \Longrightarrow N \{0, V(t)\},$$

where $V(t) = \{v_{lk, l'k'}(t)\}$ with $q = \frac{d(d+1)}{2}$ columns and rows, where

$$v_{lk, l'k'}(t) = \int K^2(u) du [\text{cov} \{\aleph_{lk}(t), \aleph_{l'k'}(t)\}]$$

and $\aleph_{lk}(t) = u_{lt} u_{kt}$.

PROOF. Before we start the proof we need to establish some notation. Let us denote by

$$W_{Tlk} = \left(\tilde{m}_l(t), \quad \tilde{m}_k(t), \quad T^{-1} \sum_{s=1}^T K_h(t-s) u_{sl} u_{sk} \right)^T$$

a 3×1 vector where $\tilde{m}_l(t) = T^{-1} \sum_{s=1}^T K_h(t-s) u_{ls}$ and $\tilde{m}_k = T^{-1} \sum_{s=1}^T K_h(t-s) u_{ks}$. Then define a $3 \times \frac{1}{2}d(d+1)$ matrix

$$W_T = \left(W_{T11}^T, W_{T12}^T, \dots, W_{T1d}^T, W_{T22}^T, \dots, W_{T2d}^T, \dots, W_{Tdd}^T \right)^T.$$

Note that $W_{Tlk} = \frac{1}{T} \sum_{s=1}^T W_{Tlks}$ and $W_T = \frac{1}{T} \sum_{s=1}^T W_{Ts}$. We shall consider the vector

$$U_{lk}(s) = (u_{ls}, u_{ks}, u_{ls}u_{ks})^T.$$

and the moments

$$M_{lk,l'k'}(s, r) = E [U_{lk}(s)U_{l'k'}^T(r)],$$

The statistics W_{Tlk} will estimate

$$\delta_{lk} = (m_l(t), \quad m_k(t), \quad \sigma_{lk}(t))^T$$

and W_T will estimate

$$\bar{\delta} = (\delta_{11}^T, \dots, \delta_{1d}^T, \delta_{22}^T, \dots, \delta_{dd}^T)^T.$$

Note that if the parametric model is correctly specified then $m_l(t) = 0$ for $l = 1, \dots, d$. Finally, let us define $Z_{Ts} = T^{-1/2}h^{1/2}(W_{Ts} - EW_{Ts})$, $Z_T = \sum_{s=1}^T Z_{Ts}$, and let

$$c = (c_{11}^T, \dots, c_{1d}^T, c_{2d}^T, \dots, c_{dd}^T)^T$$

be a $3 \times \frac{1}{2}d(d+1)$ matrix of constants composed from the 3×1 vectors

$$c_{lk} = (c_{1lk}, c_{2lk}, c_{3lk})^T.$$

In order to prove Lemma 1, we will first show that

$$\sqrt{Th}[W_T - EW_T] \implies N(0, M) \tag{A.1}$$

where

$$\text{var}(Z_T) = M. \tag{A.2}$$

We show (A.2) by noting that

$$\text{var}(c^T Z_T) = c^T M c + O(h). \tag{A.3}$$

This can be seen as follows.

$$\text{var}(c^T Z_T) = \frac{h}{T} \sum_{s=1}^T \text{var}(c^T W_{Ts}) + \frac{h}{T} \sum_{s \leq r} \text{cov}(c^T W_{Ts}, c^T W_{Tr}), \tag{A.4}$$

where,

$$\text{var} \left(c^T W_{T_s} \right) = \sum_{l \leq k} \sum_{l' \leq k'} \sum_{l'' \leq k''} c_{l''}^T \text{cov} [W_{Tlks}, W_{Tl'k's}] c_{l''k'},$$

and

$$\text{cov} \left(c^T W_{T_s}, c^T W_{T_r} \right) = \sum_{l \leq k} \sum_{l' \leq k'} \sum_{l'' \leq k''} c_{l''}^T \text{cov} [W_{Tlks}, W_{Tl'k'r}] c_{l''k'}. \quad (\text{A.5})$$

Now,

$$\text{cov} [W_{Tlks}, W_{Tl'k'r}] = K_h(t-s) K_h(t-r) M_{lk, l'k'}(s, r). \quad (\text{A.6})$$

Then,

$$\text{var} \left(c^T Z_T \right) = \sum_{l \leq k} \sum_{l' \leq k'} \sum_{l'' \leq k''} c_{l''}^T \left[\frac{h}{T} \sum_{s=1}^T K_h(t-s)^2 M_{lk, l'k'}(s, s) \right] c_{l''k'} \quad (\text{A.7})$$

and

$$\text{cov} \left(c^T W_{T_s}, c^T W_{T_t} \right) = \sum_{l \leq k} \sum_{l' \leq k'} \sum_{l'' \leq k''} c_{l''}^T \left[\frac{h}{T} \sum_{s \leq r} K_h(t-s) K_h(t-r) M_{lk, l'k'}(s, r) \right] c_{l''k'}. \quad (\text{A.8})$$

Under assumptions (H.2) to (H.4) and (H.8) to (H.10) then (A.7) is equal to

$$c^T M(t)c + O(h) \quad (\text{A.9})$$

as $T \rightarrow \infty$ and $Th \rightarrow \infty$. Where $M(t) = \left(\int_0^1 K^2(u) du M_{lk, l'k'}(t, t) \right)$, $l \leq k$, $l' \leq k'$.

Moreover, assuming (H.1) and (H.5) and Davydov's inequality (Davydov, 1967) (A.8) is of smaller order than (A.7) and therefore (A.3) holds.

In order to show (A.1) we claim that

$$\sum_{s=1}^T E |c^T Z_{T_s}|^3 \rightarrow 0 \quad (\text{A.10})$$

as T tends to infinity. Then applying the Berry-Esseen theorem for martingale differences (p. 319 of Chow and Teicher, 1980) to the triangular sequence $\{c^T Z_{ts}\}_{s=1}^T$ we obtain (A.1).

(A.10) is proved as follows.

$$\sum_{s=1}^T E |c^T Z_{T_s}|^3 = \frac{h^{3/2}}{T^{3/2}} \sum_{s=1}^T E \left| c^T (W_{T_s} - E(W_{T_s})) \right|^3$$

$$\leq \frac{h^{3/2}}{T^{3/2}} \sum_{s=1}^T \sum_{l \leq k} E \left| c_{lk}^T (W_{Tlks} - E(W_{Tlks})) \right|^3.$$

Under assumptions (H.3) and (H.7) to (H.10) it is easy to show that

$$\max_{l,k} \max_s E \left| c_{lk}^T (W_{Tlks} - E(W_{Tlks})) \right|^3 = O(h^{-3})$$

and then

$$\sum_{s=1}^T E |c^T Z_{Ts}|^3 = O((Th^3)^{-1/2})$$

that tends to zero as $Th^3 \rightarrow \infty$.

Once we have shown (A.1) we prove that

$$E(W_T) = \delta + O(h^2). \tag{A.11}$$

But then

$$\sqrt{Th} [E(W_T) - \theta] = O(Th^5)$$

that tends to zero as Th^5 tends to 0.

Finally, consider the transformation defined by

$$H(\bar{y}) = (H_{11}, H_{12}, \dots, H_{1d}, H_{22}, \dots, H_{2d}, \dots, H_{dd})^T(\bar{y})$$

where

$$H_{lk}(\bar{y}) = y_{3lk} - y_{1lk}y_{2lk}$$

and

$$\bar{y} = (y_{111}, \dots, y_{311}, y_{112}, \dots, y_{312}, \dots, y_{1dd}, \dots, y_{3dd}).$$

Apply the Mann-Wald theorem to $H(\bullet)$ (p.122 Serfling, 1980), then

$$\sqrt{Th} [H(W_T) - H(\bar{\delta})] \implies N \left(0, \frac{\partial}{\partial \delta} H(\bar{\delta}) M(t) \frac{\partial}{\partial \delta} H(\bar{\delta})^T \right).$$

It is straightforward to show that

$$\begin{aligned} H(W_T) &= \text{vech}(\widetilde{\Sigma}(t)) \\ H(\delta) &= \text{vech}(\Sigma(t)), \end{aligned}$$

and

$$V(t) = \frac{\partial}{\partial \bar{\delta}} H(\bar{\delta}) M(t) \frac{\partial}{\partial \bar{\delta}} H(\bar{\delta})^T.$$

This closes the proof of Lemma 1. ■

LEMMA 2. Assume (H.6) to (H.8). Let

$$\text{vech}[\hat{\Sigma}(t)] = T^{-1} \sum_{s=1}^T K_h(t-s) \text{vech}[\hat{u}_s \hat{u}_s^T] - \text{vech}[\hat{m}(t) \hat{m}(t)^T]$$

where $\hat{m}(t) = T^{-1} \sum_{s=1}^T K_h(t-s) \hat{u}_s$ and $\hat{u}_s = y_s - g(y_s^*, x_s, \hat{\theta}_T)$ for $t = 1, \dots, T$. Then,

$$\text{vech}[\hat{\Sigma}(t)] - \text{vech}[\tilde{\Sigma}(t)] = O_p\left(\frac{1}{Th}\right)$$

as T tends to infinity, uniformly in t .

PROOF. We have

$$\text{vech}[\hat{\Sigma}(t)] - \text{vech}[\tilde{\Sigma}(t)] = \frac{1}{T} \sum_{s=1}^T K_h(t-s) \text{vech}[\hat{u}_s \hat{u}_s^T - u_s u_s^T] - \text{vech}[\hat{m}(t) \hat{m}(t)^T - \tilde{m}(t) \tilde{m}(t)^T].$$

First we claim that

$$\frac{1}{T} \sum_{s=1}^T K_h(t-s) \text{vech}[\hat{u}_s \hat{u}_s^T - u_s u_s^T] = O_p\left(\frac{1}{Th}\right) \quad (\text{A.12})$$

This can be seen as follows. By the mean value theorem and assumption (H.7) the left hand side of equation (A.12) is equal to

$$\frac{1}{T} \sum_{s=1}^T K_h(t-s) \text{vech} \left[\left(u_s + \frac{\partial}{\partial \theta} u_s(\theta^*) (\hat{\theta}_T - \theta_0) \right) \left(u_s + \frac{\partial}{\partial \theta} u_s(\theta^*) (\hat{\theta}_T - \theta_0) \right)^T - u_s u_s^T \right] \quad (\text{A.13})$$

where θ^* lies on the line segment joining $\hat{\theta}_T$ and θ_0 and u_s are random variables that satisfy $u_s = u_s(\theta_0)$ for some θ_0 . Going through some calculations equation (A.13) can be written as

$$\frac{1}{T} \sum_{s=1}^T K_h(t-s) \text{vech} \left[\frac{\partial}{\partial \theta} u_s(\theta^*) (\hat{\theta}_T - \theta_0) (\hat{\theta}_T - \theta_0)^T \frac{\partial}{\partial \theta} u_s(\theta^*)^T \right]$$

$$\begin{aligned}
& +2\frac{1}{T}\sum_{s=1}^TK_h(t-s)\text{vech}\left[u_s(\hat{\theta}_T-\theta_0)^T\frac{\partial}{\partial\theta}u_s(\theta^*)^T\right] \\
& =\frac{1}{T}\sum_{s=1}^TK_h(t-s)\left[\frac{\partial}{\partial\theta}u_s(\theta^*)\otimes\frac{\partial}{\partial\theta}u_s(\theta^*)\right]\text{vech}\left[(\hat{\theta}_T-\theta_0)(\hat{\theta}_T-\theta_0)^T\right] \\
& \quad +\frac{1}{T}\sum_{s=1}^TK_h(t-s)\left(\frac{\partial}{\partial\theta}u_s(\theta^*)\otimes I\right)\text{vech}\left(u_s(\hat{\theta}_T-\theta_0)^T\right)
\end{aligned}$$

By assumptions (H.7) to (H.8) and Markov inequality then

$$\sup_t\left\|\frac{1}{T}\sum_{s=1}^T\left[\frac{\partial}{\partial\theta}u_s(\theta^*)\otimes\frac{\partial}{\partial\theta}u_s(\theta^*)\right]K_h(t-s)\right\|\leq\frac{1}{T}\sum_{s=1}^TE\sup_{\theta\in\Theta}\left\|\left[\frac{\partial}{\partial\theta}u_s(\theta^*)\otimes\frac{\partial}{\partial\theta}u_s(\theta^*)\right]\right\|\sup_t|K(t)|.$$

and the last expression is $O_p(\frac{1}{h})$.

$$\sup_t\left\|\frac{1}{T}\sum_{s=1}^T\left[\frac{\partial}{\partial\theta}u_s(\theta^*)\otimes I\right]K_h(t-s)\right\|\leq\frac{1}{T}\sum_{s=1}^TE\sup_{\theta\in\Theta}\left\|\left[\frac{\partial}{\partial\theta}u_s(\theta^*)\otimes I\right]\right\|\sup_t|K(t)|=O_p(\frac{1}{h}).$$

which proves (A.12).

Now, by the mean value theorem and assumption (H.7)

$$\widehat{m}(t)=\widetilde{m}(t)+\frac{1}{T}\sum_{s=1}^TK_h(t-s)\left[\frac{\partial}{\partial\theta}u_s(\theta^*)(\hat{\theta}_T-\theta_0)\right],$$

and therefore, under assumptions (H.7) and (H.8)

$$\widehat{m}(t)=\widetilde{m}(t)+O_p(T^{-1/2}h^{-1})$$

uniformly in t . Moreover, under the conditions stated in the lemma, $\widetilde{m}(t)=O_p(T^{-1/2}h^{-1/2})$, uniformly in t , and the proof of the lemma is completed. ■

PROOF OF PROPOSITION 1. Let

$$\begin{aligned}
(Th)^{1/2}\left\{\text{vech}\left[\widehat{\Sigma}(t)\right]-\text{vech}\left[\Sigma(t)\right]\right\}&=(Th)^{1/2}\left\{\text{vech}\left[\widehat{\Sigma}(t)\right]-\text{vech}\left[\widetilde{\Sigma}(t)\right]\right\}+ \\
&\quad (Th)^{1/2}\left\{\text{vech}\left[\widetilde{\Sigma}(t)\right]-\text{vech}\left[\Sigma(t)\right]\right\}
\end{aligned}$$

Apply both previous assumptions and lemmas and the proposition follows immediately. ■

PROOF OF PROPOSITION 2 Note that

$$\widehat{D}(\lambda) = (\lambda - \widehat{\lambda})\widehat{D}'(\tilde{\lambda})$$

where $\tilde{\lambda}$ lies between λ and $\widehat{\lambda}$. This implies along with the assumption of positive eigenvalues of $\Sigma(t)$ that

$$\lambda - \widehat{\lambda} = \frac{\widehat{D}(\lambda)}{D'(\lambda)} \left\{ 1 + \frac{D'(\lambda) - \widehat{D}'(\tilde{\lambda})}{\widehat{D}'(\lambda)} \right\}$$

By Proposition 1, $\widehat{\Sigma}(t) \rightarrow_p \Sigma(t)$ and then the Wielandt-Hoffman theorem (Wilkinson, 1965) implies that every λ is a continuous function of $\Sigma(t)$ and so $\widehat{\lambda} \rightarrow_p \lambda$. For all the eigenvalues. But these consistency results imply that

$$\lambda - \widehat{\lambda} = \frac{\widehat{D}(\lambda)}{D'(\lambda)} + o_p(1)$$

By Proposition 1,

$$(Th)^{1/2} \left\{ \text{vech}\widehat{\Sigma}(t) - \text{vech}\Sigma(t) \right\} \Longrightarrow N \{0, V(t)\},$$

as T tends to infinity. Then since $D(\lambda)$ is a differentiable function in λ the proposition follows immediately from Theorem A, p. 122 of Serfling (1980). ■

ACKNOWLEDGEMENTS

Financial support from the SFB 373, the National Science Foundation, the North Atlantic Treaty Organization and the *Direccion General de Enseñanza Superior*, research project PB96-1469-C05-03 is gratefully acknowledged.

References

- [1] ANDERSON, T.W. (1984): *An introduction to multivariate statistical analysis*. John Wiley, New York.
- [2] ANDREWS, D.W.K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, **11**, 560–596.
- [3] BOLLERSLEV, T, R.F. ENGLE AND D.B. NELSON (1994). ARCH Models. Ch. 45 in: R.F. Engle and D. L. McFadden, eds., *The Handbook of Econometrics*, Vol 4, Amsterdam, North-Holland.
- [4] CHOW, Y.S. AND H. TEICHER (1988). *Probability Theory: independence, interchangeability, martingales*. Springer-Verlag. New York.
- [5] DAVYDOV, Y.A. (1968). Convergence of Distributions Generated by Stationary Stochastic Processes. *Theor. Probab. Appl.*, Vol. 13, 691–696.
- [6] DONALD, S.G. (1995). Inference concerning the number of factors in a multivariate nonparametric relationship.
- [7] ENGLE R. F. (1982): “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, **50**, 987-1008.
- [8] FAIR, R.C. (1994). *Testing Macroeconometric Models*. Harvard University Press.
- [9] HÄRDLE, W. AND O. LINTON (1994). Applied Nonparametric Methods. Ch. 45 in: R.F. Engle and D. L. McFadden, eds., *The Handbook of Econometrics*, Vol 4, Amsterdam, North-Holland.
- [10] HÄRDLE, W. AND U. PARK, B. (1992). Testing Increasing Dispersion. CORE Discussion Paper, 9224.
- [11] ROBINSON, P. M. (1987): ”Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown form.” *Econometrica* 56, 875-891.
- [12] SERFLING, R.J. (1980). *Approximation theorems of mathematical statistics*, New York: John Wiley.
- [13] WILKINSON, J.H. (1965). *The Algebraic Eigenvalue Problem*. London: Oxford University Press.

7 Figures

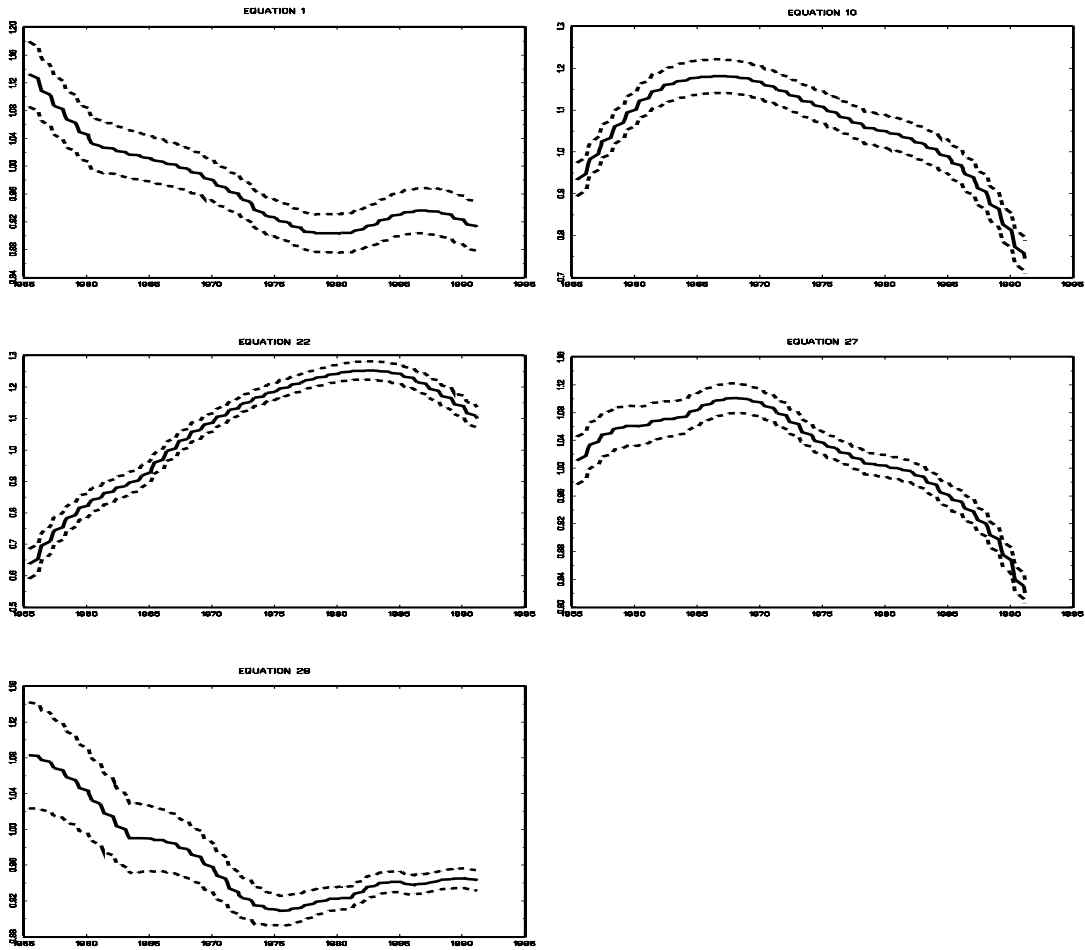


Figure 1 Conditional variances as a function of a time index. Equations 1, 10, 22, 27 and 28.

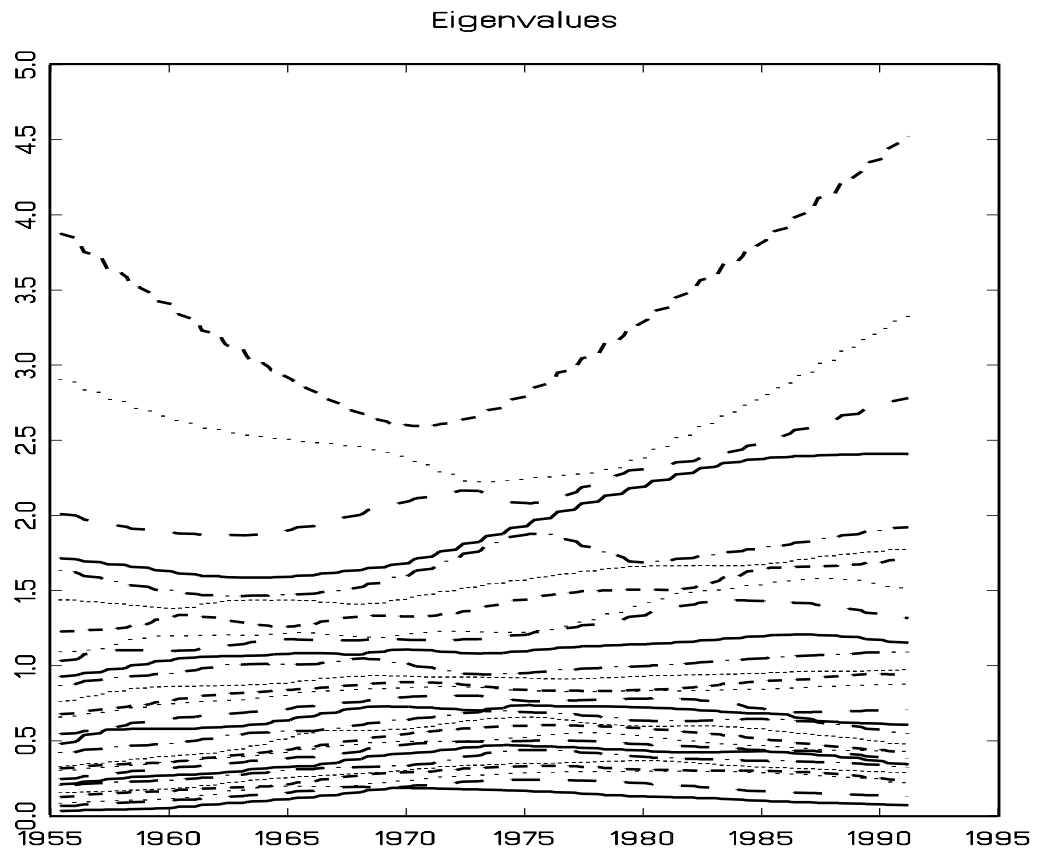


Figure 2 Eigenvalues of the covariance matrix.

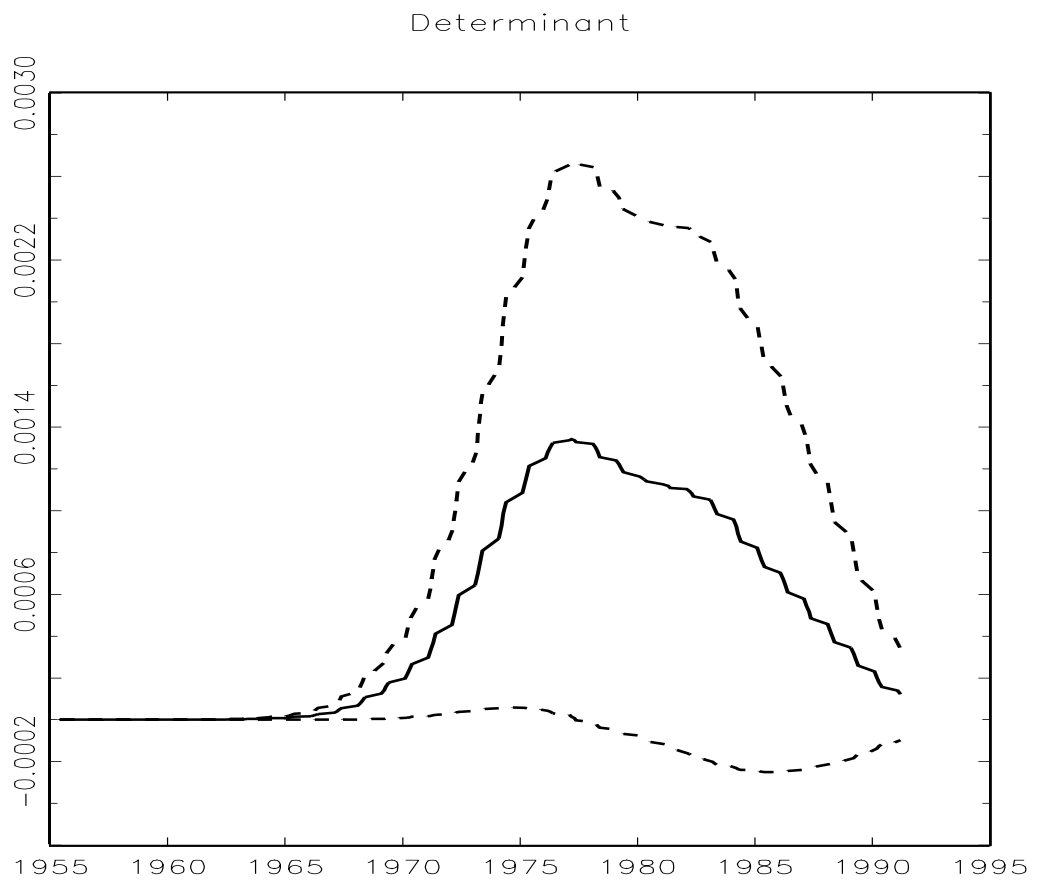


Figure 3 Determinant the covariance matrix.

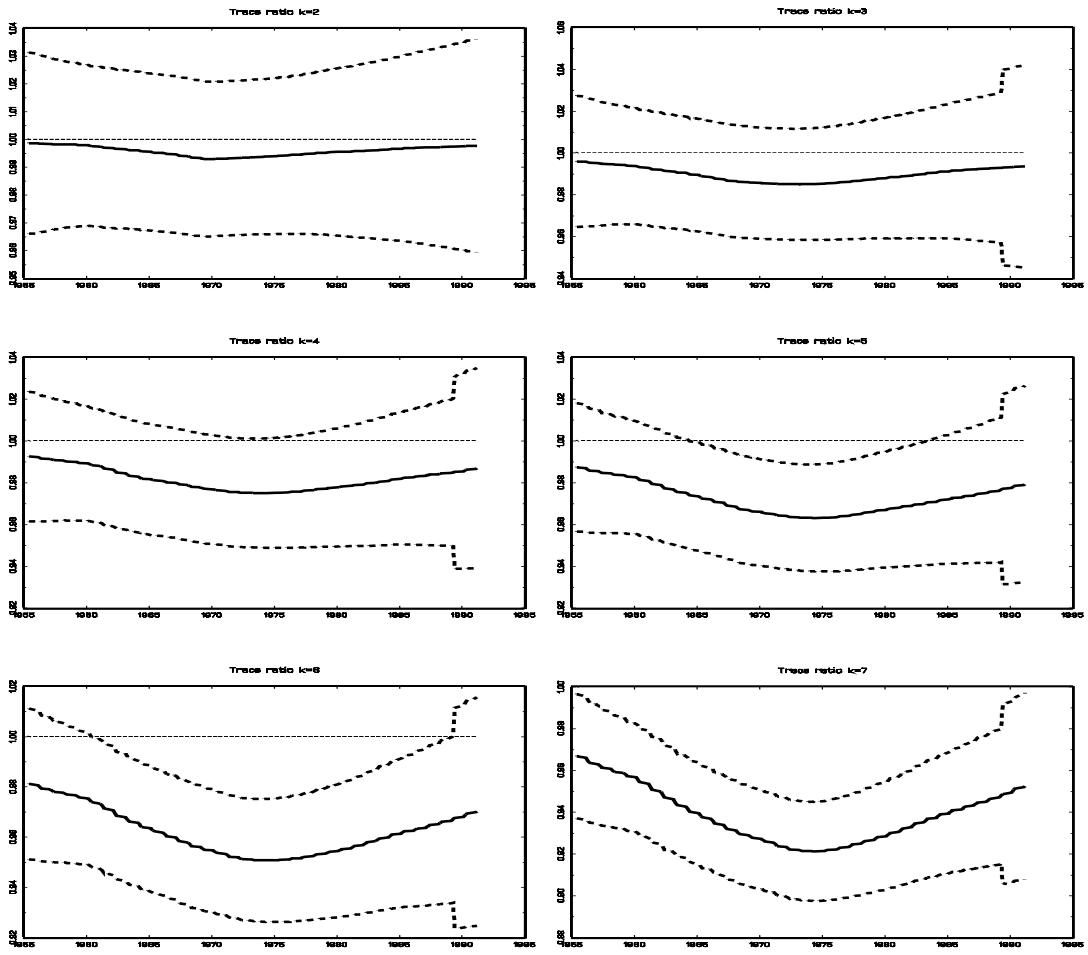


Figure 4 Trace ratio of the covariance matrix. $p = 2, 3, 4, 5, 6$ and 7 .

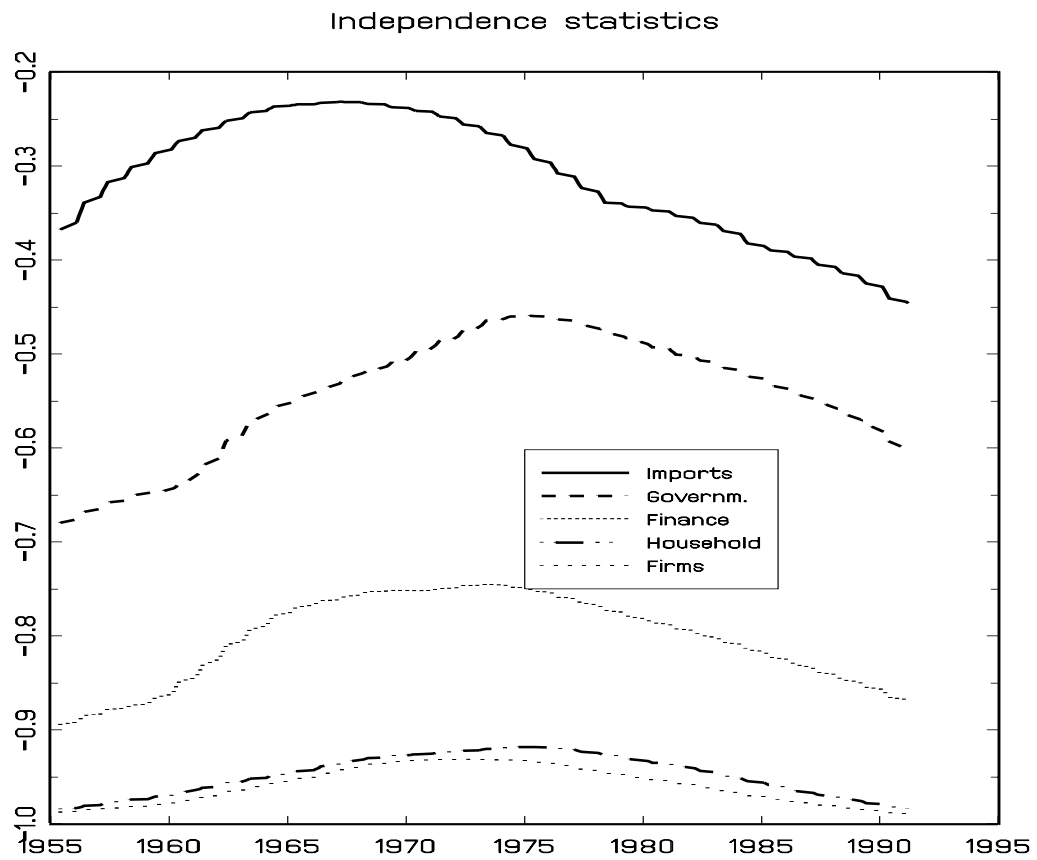


Figure 5 Statistics of independence.