



Research Report  
Statistical Research Unit  
Göteborg University  
Sweden

---

## Testing for multivariate autocorrelation

Thomas Holgersson

Research Report 2003:2  
ISSN 0349-8034

---

Mailing address:	Fax	Phone	Home Page:
Statistical Research Unit	Nat: 031-773 12 74	Nat: 031-773 10 00	<a href="http://www.stat.gu.se/stat">http://www.stat.gu.se/stat</a>
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg Sweden			

# Testing for Multivariate Autocorrelation

By

H.E.T. Holgersson

Department of Statistics

Göteborg University

SE-405 30 Göteborg

Sweden

**Abstract:** This paper concerns the problem of assessing autocorrelation of multivariate (i.e. systemwise) models. It is well known that systemwise diagnostic tests for autocorrelation often suffers from poor small sample properties in the sense that the true size overstates the nominal size gravely. The failure of keeping control of the size usually stems from the fact that the critical values (used to decide the rejection area) originate from the slowly converging asymptotic null distribution. Another drawback of existing tests is that the power may be rather low if the deviation from the null is not symmetrical over the marginal models. In this paper we consider four quite different test techniques for autocorrelation. These are (i) Pillai's trace, (ii) Roy's largest root, (iii) the maximum  $F$ -statistic and (iv) the maximum  $t^2$  test. We show how to obtain control of the size of the tests, and then examine the true (small sample) size and power properties by means of Monte Carlo simulations.

*Keywords: Autocorrelation test, Multivariate analysis, linear hypothesis, residuals.*

JEL Classification: C 32.

## I. Introduction

The history of autocorrelation testing dates back to the early 20<sup>th</sup> century, though the first test that gained popularity among practitioners is probably that of Durbin and Watson (1950). Since then, a variety of autocorrelation tests were proposed, most of which in the time domain. Some examples are the test suggested by Ljung and Box (1978) (model independent tests) and the BG test of Breusch (1978) and Godfrey (1978) (model based test). These are tests for autocorrelated errors of univariate models, but some results are available for multivariate models as well. Hosking (1980), for example, proposed multivariate versions of the model independent tests while Edgerton and Shukur (1999) proposed Multivariate versions of the BG test.

The above mentioned tests are all consistent against deviations from the null, and they limit their nominal size asymptotically. However, the multivariate testing approach has shown to behave poorly in small samples and high-dimensional models. In particular, the sizes of these tests are often so far from their nominal sizes that the power lacks meaning. For example, Edgerton and Shukur (1999) examine, among other things, the size of a Wald test for five-dimensional data with 25 observations, and report that the true size is 85% under the null hypothesis of 5% nominal size. Obviously, such a test is useless in small samples.

The failure of keeping control of the size stems from the fact that the critical values (used to decide the rejection area) originate from the slowly converging zeroeth order approximation to the true null distribution. This problem may often be corrected for, as shown in this paper, by using better approximations or tabulated critical values, and so the problem of keeping control of the size may be overcome. An additional problem with these multivariate tests is that their powers may be rather low. One reason for this is that classical multivariate tests put equal weight on each marginal model and, within each model, equal weight on each scalar parameter. Hence, if the deviation from the null is only due to one marginal model or even one single element of the parameter matrix, the power may be quite poor. So if no prior information of the type of deviation from the null is available one might want to consider some alternative approaches.

In this paper we propose four test techniques for autocorrelated errors. They are based on (i) Pillai's trace, (ii) Roy's largest root (iii), the largest  $F$ -statistic (i.e. the maximum of all marginal  $F$ -statistics); and finally (iv) the largest  $t^2$  (i.e. the maximum of all squared elementwise  $t$ -statistics). The first of these statistics is based on the sum of its arguments, while the other three use the maximum of their arguments. They represent quite different types of tests and so their optimum properties are expected to differ. This paper is concerned with this matter.

As we want our proposed testing techniques to be as general as possible, they are applied on Seemingly Unrelated Regression (SUR) models. The SUR model is a generalisation of multivariate regression with a wide field of applications (see e.g. Bewley (1986)) which in turn is a special case of univariate multiple regression and is hence quite general.

The tests are based on a two-step procedure where the first step consists of estimating the regression parameters of the SUR model and then, secondly, the autocorrelation test is performed on a secondary (multivariate) regression model based on residuals resulting from the primary regression.

In this paper we will examine the size and power properties of our proposed tests. In particular, we are interested in whether the tabulated critical values, designed for observable variables, still provide correct size when applied on residuals. Since the finite-sample null and non-null distributions of the four test statistics are unknown we will examine the properties of the tests by means of Monte Carlo simulations.

The paper is organised as follows: The next section presents some simple tests for linear restrictions of multivariate models. In section III we show how the tests, presented in Section II, can be used to test for autocorrelation by using residuals, while we present some size and power simulations in Section IV. Finally, a short summary is given in Section V.

## II. Testing linear restrictions

We previously mentioned some different methods to test for autocorrelation. In this section we explain them in more detail. The model considered in this paper is the classical SUR model

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & & \\ \vdots & & \ddots & \\ \mathbf{0} & & & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}, \text{ shortly,}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.1)$$

One crucial assumption of (2.1) is that  $Plim(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}/T) = \mathbf{Q}$ , where  $\mathbf{Q}$  is a fixed finite matrix and  $\boldsymbol{\Omega}_{(MT \times MT)} := E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']$  where  $T$  is the number of observations. It is possible to relax the assumption regarding  $\mathbf{Q}$  slightly, though it suffices well for our purposes. The model (2.1) is a generalization of multivariate regression (i.e. when  $\mathbf{X}_1 = \mathbf{X}_2 = \dots = \mathbf{X}_M$ ) with a wide field of applications (see e.g. Bewley (1986)), which in turn is a generalization of univariate multiple regression and hence quite general. The random variable of interest in this paper is the disturbance vector  $\boldsymbol{\varepsilon}$ . It is convenient to re-express the  $(MT \times 1)$  disturbances vector of (2.1) as a  $(M \times T)$  matrix,  $\boldsymbol{\varepsilon}_{(M \times T)} = [\boldsymbol{\varepsilon}_1 \cdots \boldsymbol{\varepsilon}_M]'$ . The  $t$ :th column of this matrix, i.e.  $\boldsymbol{\varepsilon}_t = [\varepsilon_{t1} \dots \varepsilon_{tM}]'$ , is assumed to follow a first order vector autoregressive process, defined by

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Gamma}\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\delta}_t, \quad (2.2)$$

where  $E[\boldsymbol{\delta}_t] = \mathbf{0}_{(M \times 1)}$ ,  $E[\boldsymbol{\delta}_t\boldsymbol{\delta}_t'] = \boldsymbol{\Sigma}_{(M \times M)}$  and  $E(\boldsymbol{\delta}_t\boldsymbol{\delta}_{t-s}') = \mathbf{0}_{(M \times M)}$  where  $m, m' = 1, \dots, M$  and  $s, t = 1, 2, \dots, T$  for  $s \neq t$ . We also assume  $\boldsymbol{\delta}_t$  to be Gaussian. In other words,  $\boldsymbol{\delta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$ . The process  $\boldsymbol{\varepsilon}_t$  is said to be stable if all eigenvalues of  $\boldsymbol{\Gamma}$  have a modulus smaller than one (Lutkepohl (1993)), which we assume holds.

Formal test statistics regarding  $\Gamma$  are frequently derived from likelihood theory, particularly through the Lagrange Multipliers. However, except in a very special case, it is impossible to concentrate any parameters out of the likelihood function of (2.1), and numerical maximization of the function with respect to all the parameters may prove to be computationally difficult (Beach and MacKinnon (1979)). But assume, for the moment, that there exist observable counterparts of  $\varepsilon_t$  and  $\varepsilon_{t-1}$  respectively,  $\mathbf{Z}_t$  and  $\mathbf{W}_t$ , say. Then (2.2) may be re-expressed as

$$\mathbf{Z}_{(M \times T)} = \Gamma_{(M \times M)} \mathbf{W}_{(M \times T)} + \boldsymbol{\delta}_{(M \times T)}, \quad (2.3)$$

with the corresponding null hypothesis of no autocorrelation

$$H_0 : \Gamma = \mathbf{0}. \quad (2.4)$$

Inferences regarding possible autocorrelation of (2.3) may then be drawn by using a classical test of linear restrictions, a well established theory, and there are several convenient procedures to test linear hypothesis regarding the parameter  $\Gamma$  available. This approach is a highly interesting candidate to likelihood-based tests. We present four test techniques below, and in section III we then show how these techniques can be used to obtain feasible tests for autocorrelation regarding (2.1).

#### i. The Pillai's trace

The usual approach to test (2.4) is to compare the restricted and unrestricted sample covariance matrixes of (2.3), i.e. the covariance matrix calculated from (2.3) under  $H_0$  and  $H_0 \cup H_A$  respectively,  $\hat{\boldsymbol{\Sigma}}_{H_0}$  and  $\hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A}$  say. Then, assuming  $\boldsymbol{\delta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma})$ , the statistic  $\mathbf{H} := T(\hat{\boldsymbol{\Sigma}}_{H_0} - \hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A})$  has a Wishart distribution (see Rao (1973) p. 549).

In addition,  $\mathbf{H}$  and  $\mathbf{E} := T \hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A}$  are independent. The matrixes  $\mathbf{H}$  and  $\mathbf{E}$  are frequently referred to as the hypothesis and error matrixes respectively. A variety of test statistics may be formulated as functions of  $\mathbf{E}$  and  $\mathbf{H}$  or in terms of the eigenvalues  $\lambda_{(M)} \geq \dots \geq \lambda_{(2)} \geq \lambda_{(1)}$  of  $\mathbf{HE}^{-1}$ . One frequently used test statistic is

$$\varphi := T \left\{ \text{tr} \left( \mathbf{H} (\mathbf{H} + \mathbf{E})^{-1} \right) \right\} = T \left\{ M - \text{tr} \left( \hat{\Sigma}_{H_0}^{-1} \hat{\Sigma}_{H_0 \cup H_A} \right) \right\} = T \sum_{m=1}^M \frac{\lambda_m}{1 + \lambda_m} \quad (2.5)$$

where  $\text{tr}(\cdot)$  is the trace operator. The statistic (2.5) is known as both the Lagrange Multiplier (LM) statistic (c.f. Bewley (1986)) and as Pillai's trace (c.f. Anderson (1984)). As  $T$  limits infinity, the statistic  $\varphi$  limits a  $\chi^2_{(\nu)}$  distribution under the null, where  $\nu$  is the total number of restriction to be tested (in our case,  $\nu = M^2$ ). However, it is well known that this asymptotic distribution does not provide control of the size in finite samples. For example, Pillai (1977) discusses the properties of several multivariate tests and their asymptotic distribution and concludes that "*the usefulness of these asymptotic results have been very little so far for the practitioner*". Higher order approximate null distributions of (2.5) have been supplied by several authors, such as Fujikoshi (1970), (1973) and Lee (1971). In particular, Muirhead (1982) has proposed an asymptotic null distribution of  $\varphi$  up to  $O(T^{-3})$ . However, this approximation involves a great deal of constants which need to be defined. A simpler approach is to use tabulated critical values instead. The (nearly exact) critical values have been calculated by several authors. A fair amount of these are reproduced in Anderson (1984), who reports a maximum difference of 3 in the third decimal place, as compared to the exact critical values. Therefore, these tabulated critical values provide a highly interesting candidate as compared to those of the limiting chi-square distribution.

## ii. Roy's largest root

A quite different statistic to that discussed above is the frequently used Roy's largest root, defined by

$$\tilde{\lambda}_{(M)} := \max_m \left\{ \lambda_m / (1 + \lambda_m) \right\}_{m=1}^M. \quad (2.6)$$

This statistic has been suggested by Roy (1945) for testing the hypothesis of (2.4). It has a rather complicated null distribution (Khatri (1972)). Asymptotic null distributions have been proposed for  $\tilde{\lambda}_{(M)}$  by Muirhead (1982), though these are of

unknown order and may not be suitable in small samples. Instead, one may use the results of Heck (1960) who proposed that

$$P\left(\hat{\lambda}_{(M)} \leq x_{\lambda}\right) \approx 1 - \alpha, \quad (2.7)$$

where  $-k \log(1 - x_{\lambda}) = z_{\lambda}$  and  $k$  is a simple function of  $T$  and  $M$ . The constant  $z_{\lambda}$  is given by Heck (1960), Table 4.1, for  $M < 6$ . The author compares the critical values of (2.7) with the critical values of the exact c.d.f. of  $\hat{\lambda}_{(M)}$ , and reports that in the most unfavourable values of  $M$ , the error is found to be five units in the fourth decimal. For  $M \geq 6$ , tabulated critical values are available in Anderson (1984), who reports similar precision. The statistic (2.7) has been frequently applied in classical multivariate analysis though it does not seem to have been appreciated in the context of diagnostic testing.

### iii. Largest marginal $F$ -statistic

When  $M$  is large, one may expect the power to detect deviations from the null to be rather low if the deviation from the null hypothesis is only in one or two of the  $M$  marginal models. One possibility to increase the power is to consider only equationwise tests and use the maximum of the marginal statistics. However, such a procedure may lead to the problem of mass significance and hence to reduction of the validity of the conclusions. Therefore we will instead consider a scale-transformed model. Assume, for the moment, that  $\Sigma$  (i.e. the covariance between the equations) is known. Then (2.3) can be transformed as  $\Sigma^{-1/2} \mathbf{Z}_i = \Sigma^{-1/2} \Gamma \mathbf{W}_i + \Sigma^{-1/2} \delta_i$ , or shortly,

$$\tilde{\mathbf{Z}}_{(M \times T)} = \tilde{\Gamma}_{(M \times M)} \mathbf{W}_{(M \times T)} + \tilde{\delta}_{(M \times T)} \quad (2.8)$$

Clearly, the marginal models of (2.8) are uncorrelated. The  $\Sigma^{-1/2}$  matrix has a (regular) inverse, namely  $\Sigma^{1/2}$ , and hence it is of full rank and therefore the columns of  $\Sigma^{-1/2}$  are linearly independent. In other words,  $\Sigma^{-1/2} \Gamma = \tilde{\Gamma} = \mathbf{0} \Leftrightarrow \Gamma = \mathbf{0}$ . The hypothesis (2.4) may then be re-expressed as



$$H_0 : (\tilde{\Gamma}_1 = \mathbf{0}) \cap \dots \cap (\tilde{\Gamma}_M = \mathbf{0}) \quad (2.9)$$

and so the  $m$ :th component of (2.9), i.e.  $H_{0m} : (\tilde{\Gamma}_m = \mathbf{0})$  where  $\tilde{\Gamma}_m$  is the  $m$ :th row-vector of  $\tilde{\Gamma}$ , may readily be tested by calculating the classical  $F$ -statistic ( $\tilde{F}_m$  say) of the  $m$ :th marginal model of (2.8). This is done by calculating the restricted and unrestricted residual covariance matrixes  $\tilde{\Sigma}_{H_0}$  and  $\tilde{\Sigma}_{H_0 \cup H_A}$ , say, from (2.8) and forming

$$\tilde{F}_m = ((T - M - 1)/M) \left( (\tilde{\sigma}_{H_0, mm} - \tilde{\sigma}_{H_0 \cup H_A, mm}) / \tilde{\sigma}_{H_0 \cup H_A, mm} \right) \quad (2.10)$$

where  $\tilde{\sigma}_{H_0, mm}$  and  $\tilde{\sigma}_{H_0 \cup H_A, mm}$  is the  $m$ :th diagonal element of  $\tilde{\Sigma}_{H_0}$  and  $\tilde{\Sigma}_{H_0 \cup H_A}$  respectively. The hypothesis  $H_{0m} : (\tilde{\Gamma}_m = \mathbf{0})$  is then rejected at the  $\alpha$  level if  $\tilde{F}_m \geq \psi$  where  $P(\tilde{F}_{(M, T-M-1)} \leq \psi) = 1 - \alpha$ , and the full hypothesis of (2.9) may then be tested by the statistic  $\tilde{F}_{max} := \max_m \{ \tilde{F}_m \}_{m=1}^M$  and it is rejected at the  $\alpha$  level if  $\tilde{F}_{max} \geq \tilde{\psi}$  where  $P(\tilde{F}_{max} \leq \tilde{\psi}) = (1 - \alpha)^{1/M}$ . Now, since  $\Sigma$  is unknown, this proposed test is not operational, but an operational, asymptotically equivalent, version may be obtained by replacing  $\Sigma$  by a consistent estimate (see Appendix A).

#### iv. Largest $t^2$ -statistic

It may well be that the deviation from the null is dominated, not by the parameters in one marginal model, but in one single element of  $\Gamma$  in (2.3), i.e. a deviation from the null where all but one of the elements of  $\Gamma$  are zero. It may then be more favourable to use the maximum of the elements as a proper test statistic. Such a statistic may be constructed as follows: Let

$$\Theta_{(M^2 \times 1)} := \sqrt{T} \mathbf{G}^{-1/2} \text{Vec}(\hat{\Gamma}) \quad (2.11)$$

where  $\hat{\Gamma}$  is the OLS estimate of the  $\Gamma$  parameter in (2.3) and  $\mathbf{G} := V[\text{Vec}(\hat{\Gamma})] = \Xi^{-1} \otimes \Sigma_{H_0 \cup H_A}$  where  $\Xi := \text{plim}(\mathbf{W}\mathbf{W}'/T)$ . Then a test regarding

the scalar element  $\Gamma_{mm'}$  is readily obtained. It may be shown that, if  $\eta$  is chosen such that  $P(\chi_{(1)}^2 \leq \eta) = (1 - \alpha)^{1/M^2}$ , then the following holds:

$$\lim_{T \rightarrow \infty} P\left(\max_i \{\Theta_i^2\}_{i=1}^{M^2} \leq \eta \mid H_0\right) = 1 - \alpha. \quad (2.12)$$

In other words, the hypothesis (2.4) is rejected at the  $\alpha$  level for  $\max_i \{\Theta_i^2\}_{i=1}^{M^2} \geq \eta$ . It may be shown that the results above still hold if the covariance matrix  $\Sigma_{H_0 \cup H_A}$  is replaced by a consistent estimate, such as  $\hat{\Sigma}_{H_0 \cup H_A} = \hat{\delta}\hat{\delta}'/T$  where  $\hat{\delta}$  is the unrestricted residuals of (2.3) (Appendix B). The exact null distribution of (2.11) is related to the regular student's  $t$  distribution, being a normal variate pre-multiplied by the inverse root of its estimated covariance matrix. Hence, a better approximation to the null distribution of  $\Theta_i^2$  may be that of a squared student's  $t$  distribution, i.e. the  $F_{(1, T-M-1)}$  distribution. Since the  $F_{(1, T-M-1)}$  distribution limits the  $\chi_{(1)}^2$  distribution asymptotically we have, analogous to (2.12),

$$\lim_{T \rightarrow \infty} P\left(\max_i \{\Theta_i^2\}_{i=1}^{M^2} \leq \tilde{\eta} \mid H_0\right) = 1 - \alpha \quad (2.13)$$

where  $P(F_{(1, T-M-1)} \leq \tilde{\eta}) = (1 - \alpha)^{1/M^2}$ . There is no guarantee that the test based on (2.13) has better small sample properties than that of (2.12). However,  $\Theta_i^2$ , is a Wald statistic (see Lutkepohl (1993) p. 93), and the Wald test based on the  $\chi^2$  distribution is well known to overestimate the nominal size (Shukur (1997)). Hence, as the  $F_{(1, \nu_2)}$  distribution has higher critical values (as compared to the  $\chi_{(1)}^2$  distribution), one may expect that the small sample size property of (2.13) is superior to that of (2.12). We will therefore determine the critical value for the maximum  $\Theta_i^2$  statistic from (2.13) rather than (2.12) and, somewhat imprecisely, refer to the test as the maximum  $t^2$  test.

Four quite different tests of the same hypothesis have been considered above. Other useful methods are the well-known tests Wilks lambda (also known as the likelihood ratio test) and Hotelling-Lawley-test (also known as the Wald test). These tests are all of the same family as they put equal weight on each marginal model and may hence be viewed as variants of Pillai's trace and will not be included here. There are, of course, numerous of other possibilities to test (2.4). Both the maximum  $F$ -test and the maximum  $t^2$ -test are examples of order statistics in its simplest form. Other more complicated linear combinations of order statistics of  $\{\Theta_i^2\}$  or  $\{F_m\}$  could be chosen as well, since (asymptotic) distributions of such statistics are usually fairly easy to obtain, see David (1969) for some examples. Furthermore, it is possible to construct a wide class of tests based on the eigenvalues  $\{\lambda_m\}$  of  $\mathbf{HE}^{-1}$ . Perlman and Olkin (1980) showed that any test with an acceptance region  $g(\lambda_1, \dots, \lambda_M) \leq c$ , where  $g$  is non-decreasing in each argument, is unbiased. They also supply monotonicity results of the power functions of such tests. In this paper, however, we will limit ourselves to the tests considered in II.i-II.iv. These tests represent distinct types and are expected to behave quite differently in small samples, and one may wonder how their power properties differ. The power properties of the Pillai's trace test and the Roy's largest root test are fairly well documented. A short review is given by Anderson (1984), who reports that *"The maximum root test has greatest power if the alternative is one-dimensional. On the other hand, if the alternative is not one-dimensional, then the maximum root test is inferior"*. Further on, none of the tests has a power function which dominates the others over the whole parameter space or even locally (Fujikoshi (1988)). In addition, the power properties of the maximum  $F$ -test and the maximum  $\Theta^2$  test are unknown. Hence, if no prior information of the type of deviation from the null is available, it may be difficult to decide which test to use. In this paper we will exemplify the possible similarities/dissimilarities among the tests. In particular, we are interested in whether some tests have good or bad over-all power properties, but we are also interested in the size properties of the tests, since the power property lacks interpretation if the true size of a test is far from its nominal size. We use Monte Carlo simulations as a tool for answering these questions. They are presented in section IV, but first we need to see how the four proposed tests may be applied in practice since we have, up to now, been assuming that the error component  $\varepsilon$  is observable, which is never true in finite samples. This matter is discussed in the following section.

### III. Testing for multivariate autocorrelation

The tests discussed in Section II concerns model (2.3) with observable variables. Since our model of interest, i.e. (2.2), consists of an unobservable variate, our above discussed tests are not feasible unless  $\varepsilon$  is replaced by some observable proxy variable, resulting in test statistics with the same asymptotical null distributions as the tests of (II.i) – (II.iv). One simple choice is the regular Feasible Generalized Least Square (FGLS) residuals,  $\hat{\varepsilon}_t := Y_t - \mathbf{X}_t \hat{\boldsymbol{\beta}}_{FGLS}$ . They are obtained by the following algorithm (c.f. Srivastava and Giles (1987)):

- i. Apply OLS to (2.1) and obtain the OLS residuals,  $\hat{\varepsilon}_{im}$  say.
- ii. Calculate  $\hat{\boldsymbol{\sigma}} := \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t' / T$  where  $\hat{\varepsilon}_t = [\hat{\varepsilon}_{t1} \dots \hat{\varepsilon}_{tM}]'$ .
- iii. Define  $\hat{\mathbf{P}} := (\hat{\boldsymbol{\sigma}}^{-1/2} \otimes \mathbf{I})$  and transform (3.1):  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \mapsto \hat{\mathbf{P}}\mathbf{Y} = \hat{\mathbf{P}}\mathbf{X}\boldsymbol{\beta} + \hat{\mathbf{P}}\varepsilon$ .
- iv. Apply OLS to  $\hat{\mathbf{P}}\mathbf{Y} = \hat{\mathbf{P}}\mathbf{X}\boldsymbol{\beta} + \hat{\mathbf{P}}\varepsilon$  and calculate the FGLS residuals

$$\hat{\varepsilon}_{(MT \times 1)} := \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{FGLS} = \mathbf{Y} - \mathbf{X} \left( (\hat{\mathbf{P}}\mathbf{X})' \hat{\mathbf{P}}\mathbf{X} \right)^{-1} (\hat{\mathbf{P}}\mathbf{X})' \hat{\mathbf{P}}\mathbf{Y}.$$

The residuals  $\hat{\varepsilon}$  are easy to calculate and play the role of a proxy variable to their unobservable counterpart  $\varepsilon$  and may be used in a straight forward manner to test the null hypothesis of zero autocorrelation. It has been shown by e.g. Schmidt (1976) that  $\hat{\varepsilon}_t \xrightarrow{\ell} \varepsilon_t$  (this derivation makes no assumption of zero autocorrelation, but implicitly assumes stationarity, which we have from our assumptions of p. 4). By using  $\hat{\varepsilon}_t$  as a proxy variable for  $\varepsilon_t$ , we may construct tests for possible autocorrelation in (2.2) by using the results of section II. As all these tests are functions of restricted and unrestricted covariance matrices (i.e.  $\hat{\boldsymbol{\Sigma}}_{H_0}$  and  $\hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A}$ ) we will start by showing how these matrices may be estimated.

An estimate of  $\boldsymbol{\Sigma}_{H_0 \cup H_A}$  (the unrestricted covariance matrix) may be obtained from the model

$$\hat{\varepsilon}_t = \boldsymbol{\Gamma} \hat{\varepsilon}_{t-1} + \boldsymbol{\delta}_t + \mathbf{v}_t \quad (3.1)$$

where  $\hat{\boldsymbol{\varepsilon}}_t$  is the  $t$ :th row of the  $\hat{\boldsymbol{\varepsilon}}_{(M \times T)}$  matrix of residuals and  $\mathbf{v}_t$  is an additive error term vanishing asymptotically. The model (3.1) depend on observable variables and play the role of an observable counterpart of (2.3). The null hypothesis of zero autocorrelation in (2.2) may then be tested by using the feasible model (3.1). Let  $\hat{\boldsymbol{\varepsilon}}_L$  be the matrix of first-order lag of  $\hat{\boldsymbol{\varepsilon}}$ ,  $\hat{\boldsymbol{\delta}}_{R,t} := \hat{\boldsymbol{\varepsilon}}_t$  and  $\hat{\boldsymbol{\delta}}_{U,t} := \hat{\boldsymbol{\varepsilon}}_t - \hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\varepsilon}}_{t-1}$  where  $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}_L' (\hat{\boldsymbol{\varepsilon}}_L\hat{\boldsymbol{\varepsilon}}_L')^{-1}$  is the OLS estimate of  $\boldsymbol{\Gamma}$ . Then we may form operational restricted and unrestricted residual covariance matrices of (3.1) by  $\hat{\boldsymbol{\Sigma}}_{H_0} := (1/T) \sum_{t=1}^T \hat{\boldsymbol{\delta}}_{R,t}\hat{\boldsymbol{\delta}}_{R,t}'$  and  $\hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A} := (1/T) \sum_{t=1}^T \hat{\boldsymbol{\delta}}_{U,t}\hat{\boldsymbol{\delta}}_{U,t}'$  respectively. From these covariance matrixes we may then calculate  $\hat{\mathbf{H}} := T(\hat{\boldsymbol{\Sigma}}_{H_0} - \hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A})$ ,  $\hat{\mathbf{E}} := T\hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A}$  and the eigenvalues  $\hat{\lambda}_{(M)} \geq \dots \geq \hat{\lambda}_{(2)} \geq \hat{\lambda}_{(1)}$  of  $\hat{\mathbf{H}}\hat{\mathbf{E}}^{-1}$ . We then have operational versions of the test statistics proposed in section II. They are defined by

$$\hat{\varphi} = \sum_{m=1}^M \frac{\hat{\lambda}_m}{1 + \hat{\lambda}_m}, \quad (3.2)$$

$$\text{and} \quad \hat{\lambda}_{(M)} = \max_m \left\{ \frac{\hat{\lambda}_m}{1 + \hat{\lambda}_m} \right\}_{m=1}^M \quad (3.3)$$

$$\hat{\Theta}_{(M^2)}^2 = \max_i \left\{ \hat{\Theta}_i^2 \right\}_{i=1}^{M^2} \quad (3.4)$$

where  $\hat{\Theta}_{(M^2 \times 1)} = \sqrt{T}\hat{\mathbf{G}}^{-1/2} \text{Vec}(\hat{\boldsymbol{\Gamma}})$ ,  $\hat{\mathbf{G}} := \hat{V}[\text{Vec}(\hat{\boldsymbol{\Gamma}})] = (\hat{\boldsymbol{\varepsilon}}_L\hat{\boldsymbol{\varepsilon}}_L'/T)^{-1} \otimes \hat{\boldsymbol{\Sigma}}_{H_0 \cup H_A}$ .

The fourth statistic may be obtained by first pre-multiplying (3.1) by  $\hat{\boldsymbol{\sigma}}^{-1/2}$  (defined on p. 11) to get  $\hat{\boldsymbol{\sigma}}^{-1/2}\hat{\boldsymbol{\varepsilon}}_t = \hat{\boldsymbol{\sigma}}^{-1/2}\boldsymbol{\Gamma}\hat{\boldsymbol{\varepsilon}}_{t-1} + \hat{\boldsymbol{\sigma}}^{-1/2}\boldsymbol{\delta}_t + \hat{\boldsymbol{\sigma}}^{-1/2}\mathbf{v}_t$ , or shortly,

$$\hat{\boldsymbol{\varepsilon}}_t = \tilde{\boldsymbol{\Gamma}}\hat{\boldsymbol{\varepsilon}}_{t-1} + \tilde{\boldsymbol{\delta}}_t + \tilde{\mathbf{v}}_t \quad (3.5)$$

Equation (3.5) is then an operational version of (2.8), with its corresponding test statistic defined by

$$\hat{F}_{(M)} = \max_m \left\{ \hat{F}_m \right\}_{m=1}^M \quad (3.6)$$

where  $\hat{F}_m = ((T - M - 1)/M) \left( \left( \hat{\sigma}_{H_0, mm} - \hat{\sigma}_{H_0 \cup H_A, mm} \right) / \hat{\sigma}_{H_0 \cup H_A, mm} \right)$  where  $\hat{\sigma}_{H_0, mm}$  and  $\hat{\sigma}_{H_0 \cup H_A, mm}$  is the  $m$ :th diagonal element of the restricted and unrestricted residuals matrices of (3.5) respectively.

The four test statistics presented above are operational versions of their counterparts in Section II, and the critical values used to test  $H_0 : \Gamma = \mathbf{0}$  are the same as those of statistics i-iv in Section II \*.

Now, there are some uncertainties related to the tests proposed above. Firstly, they are based on proxy variables to the unobservable error variables, whereas the critical values are designed for observable variables, and this may have an impact on the size properties. Secondly, they are expected to have different optimality properties and one may wonder if and how the power properties differ among the four tests. These size and power properties are not possible to derive analytically, as the null and non null distributions of the statistics are unknown in finite samples. Therefore, we will use Monte Carlo simulations in order to shed some light on these questions. Details on the simulations are presented and discussed in the following section.

\* The tabulated critical values do not cover all possible sample sizes. Typically, they are tabulated only for sample sizes  $T = 10, 20, \dots$  or similar intervals. In order to obtain critical values for arbitrary sample sizes, we have used linear regression to interpolate critical values. Details are supplied in Appendix C.

#### IV. Size and power

We previously discussed the fact that the test statistics presented in Section II are constructed for observable variables rather than for residuals, and this may have an impact on the finite-sample size properties. Also, the tests based on the largest  $\Theta^2$  statistic and the largest  $F$  statistic rely on asymptotic properties. We therefore need to examine their finite sample properties by Monte Carlo simulations in order to obtain an idea of the agreement of the tests' true sizes to the nominal sizes. Further on, one of our main purposes in this paper is to examine the possible differences of the powers of our proposed test techniques. In particular, we would like to see if some of the tests have superior over-all properties (i.e. have high power against a wide class of deviations from the null). In order to perform these examinations, we need to specify some models for the so-called primary regression used to generate the residuals on which we form the secondary regression, used to perform the autocorrelation tests. We will also specify some null and non-null parameterisations of the residual process for the power simulations. Below we present the models used in the simulations.

**Parameters used in the simulated models.** The residuals proposed in section III do not depend on the regression parameters. However, they do depend on the properties of the regressors as well as on the error covariance matrix. Below, in Tables I and II, we display the choices used in our simulations.

**Table I. Factors that are used in the size and power simulations.**

Factor	7-dimensional data	3-dimensional data
Error Covariance matrix, $\Sigma$	$toeplitz(4 \ 2 \ 1 \ .5 \ .2 \ 0 \ 0)^*$	$toeplitz(4 \ 2 \ 1)^*$
Number of regressors.	4	4
Number of observations, $T$ .	25-120	25-120
Distribution of regressors.	$X_{it} \stackrel{iid}{\sim} U[0,1]$	$X_{it} \stackrel{iid}{\sim} U[0,1]$
Distributions of errors.	$\delta_t \stackrel{iid}{\sim} N[\mathbf{0}, \Sigma]$	$\delta_t \stackrel{iid}{\sim} N[\mathbf{0}, \Sigma]$

**Table II. Parameters of model (2.2) used in power simulations.**

$\Gamma$	7-dimensional data	3-dimensional data
Power model 1	$\text{toeplitz}(.28 \ .2 \ .1 \ .05 \ .02 \ 0 \ 0)^*$	$\text{toeplitz}(.28 \ .2 \ .1)^*$
Power model 2	$\text{toeplitz}(.6 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^*$	$\text{toeplitz}(.6 \ 0 \ 0)^*$
Power model 3	$\Gamma_2 = [.35 \ .2 \ .15 \ .1 \ .05 \ .04 \ 0]$ all other $\Gamma_{ij} = 0$ .	$\Gamma_2 = [.35 \ .2 \ .15]$ all other $\Gamma_{ij} = 0$ .
Power model 4	$\Gamma_{2,2} = 0.7$ all other $\Gamma_{ij} = 0$ .	$\Gamma_{2,2} = 0.7$ . all other $\Gamma_{ij} = 0$ .

\* The toeplitz operator provides a convenient technique for choosing the elements in high-dimensional matrix-valued parameters. The upper triangle of the symmetric matrix  $\mathbf{A} = \text{toeplitz}(\psi_1, \dots, \psi_M)$  is defined by its diagonal elements  $A_{mm} = \psi_1$ ,  $m = 1, \dots, M$ , the first off-diagonal by  $A_{m(m+1)} = \psi_2$ ,  $m = 1, \dots, (M-1)$ , the second off-diagonal by  $A_{m(m+2)} = \psi_3$ ,  $m = 1, \dots, (M-2)$  and so on.

The parameters in Table II represent four different power parameterizations, one where (nearly) all elements of  $\Gamma$  deviate from zero (model 1), one where only the diagonal elements of  $\Gamma$  deviate from zero (model 2), one where only one marginal model deviates from zero (model 3) and one model where only a single element of  $\Gamma$  deviates from zero (model 4).

When considering table III, we see that the Pillai's trace, the largest root test and the maximum  $F$  test hold control of the size even in the smallest sample size, staying within our defined acceptable range of  $\pm 1\%$  from its nominal size. Unfortunately, the maximum  $t^2$  test does not quite control the size, it needs about 50 observations before it is reasonable. Hence the power properties of the maximum  $t^2$  test should be interpreted with care in small samples. Moving on to Table IV, the power properties of the tests on the 3-dimensional data do not differ markedly. However, the power differs up to a good 10% in the 7-dimensional process. The largest root test has higher power than the Pillai's trace test over the whole range of sample sizes.



Nobs	Pillaitrace	Largest root	Max t2	MaxF	Pillaitrace	Largest root	Maxt2	MaxF
25	.044	.050	.035	.042	.048	.058	.116	.049
30	.045	.052	.035	.047	.050	.062	.079	.045
35	.045	.050	.039	.048	.047	.056	.064	.049
40	.047	.052	.039	.045	.047	.058	.059	.050
45	.048	.052	.040	.046	.044	.052	.056	.047
50	.048	.051	.040	.047	.045	.051	.046	.043
60	.048	.050	.044	.046	.050	.049	.048	.047
80	.048	.050	.044	.050	.045	.046	.048	.050
100	.048	.051	.045	.048	.046	.048	.043	.048
120	.048	.050	.049	.049	.048	.058	.047	.051

Nobs	Pillaitrace	Largest root	Max t2	MaxF	Pillaitrace	Largest root	Maxt2	MaxF
25	.274	.321	.193	.282	.281	.307	.336	.387
30	.414	.464	.289	.399	.446	.549	.409	.536
35	.543	.586	.394	.503	.629	.737	.553	.674
40	.662	.699	.491	.605	.782	.857	.636	.780
45	.742	.778	.575	.687	.893	.929	.717	.857
50	.820	.844	.660	.754	.954	.966	.809	.909
60	.911	.923	.778	.859	.994	.994	.909	.969
80	.984	.986	.919	.962	1	1	.987	.997
100	.998	.997	.977	.991	1	1	.998	1
120	1	1	.994	.998	1	1	1	1

Nobs	Pillaitrace	Largest root	Max t2	MaxF	Pillaitrace	Largest root	Maxt2	MaxF
25	.631	.535	.398	.509	.521	.309	.397	.509
30	.835	.729	.593	.707	.791	.549	.513	.707
35	.935	.859	.761	.841	.945	.756	.658	.844
40	.981	.940	.871	.928	.992	.884	.792	.927
45	.995	.973	.937	.972	.999	.952	.890	.968
50	.998	.988	.965	.986	1	.984	.947	.988
60	1	.999	.994	.998	1	.999	.992	.998
80	1	1	1	1	1	1	.999	1
100	1	1	1	1	1	1	1	1
120	1	1	1	1	1	1	1	1

Nobs	Pillaitrace	Largest root	Max t2	MaxF	Pillaitrace	Largest root	Maxt2	MaxF
25	.233	.272	.144	.215	.102	.100	.152	.100
30	.337	.386	.210	.316	.131	.157	.119	.140
35	.443	.498	.281	.429	.177	.216	.129	.212
40	.556	.610	.367	.531	.218	.290	.156	.292
45	.651	.703	.451	.630	.271	.365	.180	.376
50	.731	.783	.533	.710	.328	.467	.226	.474
60	.853	.888	.663	.837	.451	.631	.331	.641
80	.961	.975	.858	.957	.694	.867	.567	.860
100	.993	.996	.945	.990	.869	.966	.769	.967
120	.999	1	.981	.998	.953	.994	.892	.993

**Table VII.** *Power for model 4, M=3.* *Power for model 4, M=7.*

Nobs	Pillaitrace	Largest root	Max t2	MaxF		Pillaitrace	Largest root	Maxt2	MaxF
25	.441	.541	.446	.467		.120	.120	.172	.147
30	.642	.725	.640	.653		.174	.225	.204	.258
35	.794	.856	.791	.793		.243	.355	.307	.403
40	.893	.929	.886	.884		.326	.488	.456	.535
45	.946	.967	.946	.939		.425	.613	.591	.658
50	.974	.985	.969	.969		.511	.724	.713	.767
60	.995	.997	.993	.993		.692	.880	.870	.891
80	1	1	.999	1		.913	.985	.982	.983
100	1	1	1	1		.989	.999	.999	.998
120	1	1	1	1		.999	1	1	1

In Table V, the opposite effect occurs. Here, the Pillai's trace test is dominating with a power difference of up to a good 10%, and the same pattern is shown in the 3- and 7-dimensional models. Further on, in Table VI, where only one marginal model deviates from the null, the maximum  $F$  test and the largest root test are dominating and the difference between the two tests is only about 5% at the most, in favor of the largest root test. The Pillai's trace test is markedly inferior to the maximum  $F$  test and the largest root test in the 7 dimensional model. Finally, in Table VII, some really interesting facts are visualized. In this parameterization, where only one single element deviates from the null, the largest root test is only slightly better than the Pillai's trace test in the 3-dimensional data. Though for the 7-dimensional data the power differs by up to 25%, in favor of the largest root-test. Hence the difference in power among the tests may be quite large in some parameterizations. Also, note that the maximum  $F$  test has a power function only slightly lower than that of the largest root test. In all, the largest root test seems to have the highest power for all models except when precisely the marginal parameters deviate from zero (i.e. model 2) where Pillai's trace test is dominating. A somewhat surprising fact though, is that the maximum  $t^2$  statistic does not dominate the power functions for any parameterization.

## V. Summary

In this paper we have proposed four simple tests against multivariate autocorrelation to be used in SUR models or multivariate regressions. Two of them, namely the Pillai's trace test and Roy's largest root test are frequently applied in multivariate analysis but do not seem to have been appreciated in diagnostic testing. One reason for this may be that the convergence rates to their asymptotic null distributions are rather slow, in particular in high dimensional data. We have here demonstrated that this size problem is readily overcome by using tabulated critical values (available in the literature). Even though these critical values are derived for observable data, we have shown in this paper that they still provide good control of the size when applied on residuals. When it comes to the other two tests, the largest  $t^2$  test and the largest  $F$  test, our simulations have shed some light on their properties as well. The largest  $t^2$  test has shown to be inferior to the other tests in terms of size. It underestimates the true size in the three-dimensional case while it overestimates in the seven-dimensional models. In contrast, the size of the largest  $F$  test stays in close proximity to the nominal size, even in high dimensional data and small samples.

Furthermore, we have demonstrated that the power of the different tests may be quite different, depending on the type of deviation from the null. In particular, in one of our power comparisons the difference between the powers of two tests was as high as 25%. Hence the choice of test may be of crucial importance. Expectedly, no single test has uniformly highest power. Nevertheless, in these simulations, we found that two tests have a good over-all power property. These are the largest root test and the largest  $F$  test. Unfortunately though, they do not have simultaneous maximum powers for the same type of deviations and so there is no obvious choice of test for a situation where one does not know much of the type of possible deviation from the null. However, in terms of simplicity, the largest  $F$  test may be preferable since it does not require tabulated critical values. Also, the power of this test is not far below the power of the largest root test in the situations where the latter is dominating. Hence, from this point of view, the largest  $F$  test may be preferable in an applied study.

Acknowledgements: The author is indebted to Eva Andersson and Ghazi Shukur for valuable comments on earlier versions of the manuscript.

## Appendix A.

In Section II.iii. we considered the asymptotic distribution of  $\tilde{F}_{\Sigma, \max} := \max_m \left\{ \tilde{F}_{\Sigma, m} \right\}_{m=1}^M$  where  $\tilde{F}_{\Sigma, m}$  is the  $F$ -statistic of the  $m$ :th marginal model of the transformed model (2.3), i.e.  $\Sigma^{-1/2} \mathbf{Z}_i = \Sigma^{-1/2} \Gamma \mathbf{W}_i + \Sigma^{-1/2} \delta_i$ , in matrix form  $\tilde{\mathbf{Z}} = \tilde{\Gamma} \mathbf{W} + \tilde{\delta}$  and the index  $\Sigma$  of  $F_{\Sigma}$  indicates that the true covariance matrix has been used. Now, consider the corresponding statistic  $\tilde{F}_{\hat{\Sigma}, \max} := \max_m \left\{ \tilde{F}_{\hat{\Sigma}, m} \right\}_{m=1}^M$  where the true covariance matrix has been replaced by an estimate such that  $\hat{\Sigma} \xrightarrow{P} \Sigma$ . Then the following holds:

**Proposition 1:**  $\tilde{F}_{\hat{\Sigma}, \max} \xrightarrow{\ell} \tilde{F}_{\Sigma, \max}.$

**Proof:** Firstly, consider the untransformed model  $\mathbf{Z} = \Gamma \mathbf{W} + \delta$  and its (OLS) residual vector  $\hat{\delta} = (\mathbf{Z} - \hat{\Gamma} \mathbf{W}) = \mathbf{Z} - \mathbf{Z} \mathbf{W}' (\mathbf{W} \mathbf{W}')^{-1} \mathbf{W}$  with corresponding residual covariance matrix  $\hat{\delta} \hat{\delta}' / T$ . For the case of restricted covariance matrix,  $\mathbf{W}$  only consists of a unit vector (corresponding to the intercept parameter) and there is hence no need to treat the restricted and unrestricted covariance matrix separately. Clearly, under the null hypothesis,  $(\hat{\delta} \hat{\delta}' / T) \xrightarrow{P} \Sigma_{H_0}$ . Hence  $(\hat{\delta} \hat{\delta}' / T) = O(1)$ , both for the restricted and unrestricted estimate. Next, consider the transformed model  $\tilde{\mathbf{Z}} = \tilde{\Gamma} \mathbf{W} + \tilde{\delta}$  and its corresponding residual vector

$$\hat{\tilde{\delta}} = (\tilde{\mathbf{Z}} - \hat{\tilde{\Gamma}} \mathbf{W}) = \tilde{\mathbf{Z}} - \tilde{\mathbf{Z}} \mathbf{W}' (\mathbf{W} \mathbf{W}')^{-1} \mathbf{W} = \Sigma^{-1/2} (\mathbf{Z} - \mathbf{Z} \mathbf{W}' (\mathbf{W} \mathbf{W}')^{-1} \mathbf{W}) = \Sigma^{-1/2} \hat{\delta}. \quad (\text{A1})$$

Hence the residual covariance matrix of the transformed model is  $(\hat{\tilde{\delta}} \hat{\tilde{\delta}}' / T) = \Sigma^{-1/2} (\hat{\delta} \hat{\delta}' / T) \Sigma^{-1/2}$  where the index  $\Sigma$  of  $\hat{\tilde{\delta}}_{\Sigma}$  indicates that the true covariance matrix  $\Sigma$  have been used in the transformation  $\mathbf{Z} \mapsto \Sigma^{-1/2} \mathbf{Z}$ . The residual covariance matrix of the feasible transformed model may then likewise be written as  $(\hat{\tilde{\delta}}_{\hat{\Sigma}} \hat{\tilde{\delta}}_{\hat{\Sigma}}' / T) = \hat{\Sigma}^{-1/2} (\hat{\delta} \hat{\delta}' / T) \hat{\Sigma}^{-1/2}$ . Now, we have by assumption  $\hat{\Sigma} \xrightarrow{P} \Sigma$  and hence

$\hat{\Sigma}^{-1/2} \xrightarrow{p} \Sigma^{-1/2}$  or equivalently,  $(\hat{\Sigma}^{-1/2} - \Sigma^{-1/2}) = o(1)$ , and so  $\hat{\Sigma}^{-1/2} = \Sigma^{-1/2} + o(1)$ . We then have

$$\begin{aligned} & \left( \hat{\delta}_{\hat{\Sigma}} \hat{\delta}'_{\hat{\Sigma}} / T \right) - \left( \hat{\delta}_{\Sigma} \hat{\delta}'_{\Sigma} / T \right) = \hat{\Sigma}^{-1/2} \left( \hat{\delta} \hat{\delta}' / T \right) \hat{\Sigma}'^{-1/2} - \Sigma^{-1/2} \left( \hat{\delta} \hat{\delta}' / T \right) \Sigma'^{-1/2} = \\ & \left( \Sigma^{-1/2} + o(1) \right) \left( \hat{\delta} \hat{\delta}' / T \right) \left( \Sigma^{-1/2} + o(1) \right)' - \left( \Sigma^{-1/2} \left( \hat{\delta} \hat{\delta}' / T \right) \Sigma'^{-1/2} \right) = \\ & o(1) \left( \hat{\delta} \hat{\delta}' / T \right) \Sigma'^{-1/2} + \Sigma^{-1/2} \left( \hat{\delta} \hat{\delta}' / T \right) o(1) + o(1) \left( \hat{\delta} \hat{\delta}' / T \right) o(1) = \\ & o(1) O(1) + O(1) o(1) + o(1) O(1) o(1) = o(1). \end{aligned} \quad (\text{A2})$$

Hence,

$$\left( \hat{\delta}_{\hat{\Sigma}} \hat{\delta}'_{\hat{\Sigma}} / T \right) - \left( \hat{\delta}_{\Sigma} \hat{\delta}'_{\Sigma} / T \right) \xrightarrow{p} \mathbf{0}, \quad (\text{A3})$$

both for the restricted and unrestricted covariance matrix (under the null). Next, define  $\mathbf{H}_{\Sigma} := \left( \hat{\delta}_{\Sigma, H_0} \hat{\delta}'_{\Sigma, H_0} / T \right) - \left( \hat{\delta}_{\Sigma, H_0 \cup A} \hat{\delta}'_{\Sigma, H_0 \cup A} / T \right)$  and  $\mathbf{E}_{\Sigma} := \left( \hat{\delta}_{\Sigma, H_0 \cup A} \hat{\delta}'_{\Sigma, H_0 \cup A} / T \right)$  where the subscripts indicate that the residuals  $\delta$  have been calculated under  $H_0$  and  $H_0 \cup H_A$  respectively, assuming  $\Sigma$  is known. Similarly,  $\mathbf{H}_{\hat{\Sigma}}$  and  $\mathbf{E}_{\hat{\Sigma}}$  are the matrices corresponding to  $\mathbf{H}_{\Sigma}$  and  $\mathbf{E}_{\Sigma}$ , though with  $\Sigma$  replaced by  $\hat{\Sigma}$ . From (A3) we have that

$$\mathbf{H}_{\hat{\Sigma}} - \mathbf{H}_{\Sigma} \xrightarrow{p} \mathbf{0}, \quad \mathbf{E}_{\hat{\Sigma}} - \mathbf{E}_{\Sigma} \xrightarrow{p} \mathbf{0}. \quad (\text{A4})$$

Further, we have, from Rao (1973) p. 458,  $F_{\Sigma, m} = \left( H_{\Sigma, mm} / \nu_1 \right) / \left( E_{\Sigma, mm} / \nu_2 \right) \sim F_{(\nu_1, \nu_2)}$ ,  $m = 1, \dots, M$ . Similarly, define  $F_{\hat{\Sigma}, m} = \left( H_{\hat{\Sigma}, mm} / \nu_1 \right) / \left( E_{\hat{\Sigma}, mm} / \nu_2 \right)$ . Then, following Rao (1973),  $X_n - Y_n \xrightarrow{p} 0$ ,  $Y_n \xrightarrow{\ell} Y$ ,  $\Rightarrow X_n \xrightarrow{\ell} Y$ , and hence as  $F_{\Sigma, m} \sim F_{(\nu_1, \nu_2)}$  using (A4) we get  $F_{m, \hat{\Sigma}} \xrightarrow{\ell} F_{(\nu_1, \nu_2)}$  and so  $\max_m \left\{ F_{m, \hat{\Sigma}} \right\}_{m=1}^M \xrightarrow{\ell} \max_m \left\{ F_{m, \Sigma} \right\}_{m=1}^M$ . In other words, we have established that the  $F$ -statistic of the  $m$ :th marginal model of the feasible transformed model (i.e.  $\mathbf{Z} \mapsto \hat{\Sigma}^{-1/2} \mathbf{Z}$ ) have asymptotically the same distribution as the  $F$ -statistic of the theoretical transformed model (i.e.  $\mathbf{Z} \mapsto \Sigma^{-1/2} \mathbf{Z}$ )  $\square$ .

**Appendix B.** In section II.iv we considered the asymptotic distribution of  $\Theta_{\Sigma, \max}^2 := \max_i \left\{ \Theta_{\Sigma, i}^2 \right\}_{i=1}^{M^2}$  where  $\Theta_{\Sigma, i}$  is the  $i$ :th element of the scaled variate  $\Theta_{\Sigma} := \sqrt{T} \mathbf{G}_{\Sigma}^{-1/2} \text{Vec}(\hat{\Gamma})$ , where  $\mathbf{G}_{\Sigma} := \Xi^{-1} \otimes \Sigma_{H_0 \cup H_A}$ ,  $\Xi := \text{plim}(\mathbf{W}\mathbf{W}'/T)$ ,  $\Sigma = \text{Cov}[\delta]$ , and  $\Xi$  and  $\Sigma$  are positive definite (p.d.) by assumption. Note that if  $\Xi$  is p.d., then so is  $\Xi^{-1}$  (Harville (1997) p. 214). Furthermore,  $\mathbf{G}_{\Sigma}$  is p.d. since  $\Xi^{-1}$  and  $\Sigma$  are p.d. (Harville (1997) p. 369). Now, consider the estimate  $\mathbf{G}_{\hat{\Sigma}} = (\mathbf{W}\mathbf{W}'/T)^{-1} \otimes \hat{\Sigma}_{H_0 \cup H_A}$  where  $\hat{\Sigma}_{H_0 \cup H_A}$  is any estimate such that  $\hat{\Sigma}_{H_0 \cup H_A} \xrightarrow{p} \Sigma_{H_0 \cup H_A}$ , and let  $\Theta_{\hat{\Sigma}} := \sqrt{T} \mathbf{G}_{\hat{\Sigma}}^{-1/2} \text{Vec}(\hat{\Gamma})$ . Further, let  $\eta$  be chosen such that  $P(\chi_{(1)}^2 \leq \eta) = (1 - \alpha)^{1/M^2}$  where  $\alpha$  is the test level desired. Then the following holds:

**Proposition 2:**  $\lim_{T \rightarrow \infty} P\left(\max_i \left\{ \Theta_{\hat{\Sigma}, i}^2 \right\}_{i=1}^{M^2} \leq \eta\right) = 1 - \alpha.$

**Proof:** Firstly, we note that, if  $h: \mathbb{R}^k \mapsto \mathbb{R}^l$  and  $g: \mathbb{R}^m \mapsto \mathbb{R}^n$  are continuous functions, and  $\mathbf{X}$  is a finite-dimension random vector, then  $\mathbf{X}_T \xrightarrow{\ell} \mathbf{X} \Rightarrow h(\mathbf{X}_T) \xrightarrow{\ell} h(\mathbf{X})$  and  $\mathbf{X}_T \xrightarrow{p} \mathbf{X} \Rightarrow g(\mathbf{X}_T) \xrightarrow{p} g(\mathbf{X})$  hold (Shao (1999) p.42). Now, it is well known that under some quite general conditions,  $\sqrt{T} \text{Vec}(\hat{\Gamma} - \Gamma) \xrightarrow{\ell} N(\mathbf{0}, \mathbf{G})$  where  $\mathbf{G} := \Xi^{-1} \otimes \Sigma$ , (Lutkepohl p. 66). Since  $\hat{\Sigma} \xrightarrow{p} \Sigma$  and  $(\mathbf{W}\mathbf{W}'/T) \xrightarrow{p} \Xi$  by assumption, it follows by the product rule that  $\mathbf{G}_{\hat{\Sigma}} = (\hat{\Xi}^{-1} \otimes \hat{\Sigma}) \xrightarrow{p} (\Xi^{-1} \otimes \Sigma) = \mathbf{G}_{\Sigma}$ . Unique continuous versions of the square root of a p.d. matrix,  $\mathbf{G}^{1/2}$  say, are the Cholesky square root and the symmetric p.d. square root (Fahrmeir and Kaufman (1985) p. 348). Hence, if we restrict the matrix root to be of this class, we have  $\mathbf{G}_{\hat{\Sigma}}^{1/2} \xrightarrow{p} \hat{\mathbf{G}}_{\Sigma}^{1/2}$ , and since the matrix inverse mapping is continuous (Davidson (1994) p. 287) we have  $\hat{\mathbf{G}}_{\Sigma}^{-1/2} \xrightarrow{p} \hat{\mathbf{G}}_{\Sigma}^{-1/2}$ . Hence, by the product rule for central limit theorems (Hamilton (1994) p. 185), we have  $\Theta_{\hat{\Sigma}} := \sqrt{T} \hat{\mathbf{G}}_{\Sigma}^{-1/2} \text{Vec}(\hat{\Gamma} - \Gamma) \xrightarrow{\ell} N(\mathbf{0}_{M^2 \times 1}, \mathbf{I}_{M^2 \times M^2})$  and so  $\Theta_{\hat{\Sigma}, i} \xrightarrow{\ell} iidN(0, 1)$ . Since

$h(X) = X^2$  is continuous, we then have  $\Theta_{\hat{\Sigma},i}^2 \xrightarrow{\ell} \chi_{(1)}^2$  where  $\Theta_{\hat{\Sigma},i}^2$  is the square of the  $i$ :th element of  $\Theta_{\hat{\Sigma}}$ . Furthermore, as  $g(\mathbf{X}) = \max(\mathbf{X})$  is continuous at  $\mathbf{X}$  (for a continuous variable  $\mathbf{X}$ ), it follows that  $\Theta_{\hat{\Sigma}}^2 \xrightarrow{\ell} \chi_{(1)}^2 \Rightarrow \max_i \{\Theta_{\hat{\Sigma},i}^2\}_{i=1}^{M^2} \xrightarrow{\ell} \max_i \{\chi_{(1),i}^2\}_{i=1}^{M^2}$  where  $\chi_{(1)}^2$  is an  $(M^2 \times 1)$  vector of *iid*  $\chi_{(1)}^2$  variates. In other words, we have established that  $\lim_{T \rightarrow \infty} P\left(\max_i \{\Theta_{\hat{\Sigma},i}^2\}_{i=1}^{M^2} \leq k\right) = P\left(\max_i \{\chi_{(1),i}^2\}_{i=1}^{M^2} \leq k\right)$  for some constant  $k$ , and since  $P\left(\max_i \{\chi_{(1-\alpha),i}^2\}_{i=1}^{M^2} \leq k\right) = \prod_{i=1}^{M^2} P\left(\chi_{(1-\alpha),i}^2 \leq k\right) = (1-\alpha)^{M^2}$ , we may choose  $\tilde{k}$  such that  $P\left(\chi_{(1-\alpha),i}^2 \leq \tilde{k}\right) = (1-\alpha)^{1/M^2}$  which completes the proof  $\square$ .

### Appendix C:

The following linear regression function was used to interpolate critical values for the Roys' largest root statistic and Pillai's trace statistic:

Let  $k_{\theta,M,T}$  be the "observation" (originating from a table) of the critical value for the statistic  $\theta$  for  $M$ -dimensional data for sample size  $T$ . Then we have used the linear regression model

$$k_{\theta,M,T} = \beta_0 + \beta_1 \ln(T) + \beta_2 (\ln(T))^2 + \varepsilon_{M,T} \quad (\text{C1})$$

and the corresponding estimated models (parameters estimated with ordinary least square), used to interpolate are;

**Roy's largest root:**  $\hat{k}_{\lambda,7,T} = 1.314 - 0.23 \ln(T) + 0.0024 (\ln(T))^2$ ,  $R_{adj.}^2 = 0.993$ .

13 "observations" were available for interpolation range  $5 \leq T \leq 100$ .

**Pillai's trace:**  $\hat{k}_{\lambda,3,T} = 2.207 + 1.378 \ln(T) - 0.142 (\ln(T))^2$ ,  $R_{adj.}^2 = 0.997$ .

10 "observations" were available for interpolation range  $14 \leq T \leq 124$ .

$\hat{k}_{\lambda,7,T} = 4.283 + 1.971 \ln(T) - 0.193 (\ln(T))^2$ ,  $R_{adj.}^2 = 0.999$ .

10 "observations" were available for interpolation range  $18 \leq T \leq 128$ .

## References.

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York, Wiley.
- Beach, C. M. and J. G. MacKinnon (1979). "Maximum Likelihood Estimation of Singular Equation Systems With Autoregressive Disturbances." *International Economic Review* 20(2): 459-464.
- Bewley, R. (1986). *Allocation models, Specification, Estimation and Applications*, Ballinger Publishing Company.
- Breusch, T. S. (1978). "Testing for Autocorrelation in Dynamic Linear Models." *Australian Economic Papers* 17: 334-355.
- David, H. A. (1969). *Order Statistics*, Wiley.
- Davidson, J. (1994). *Stochastic Limit Theory*. New York, Oxford University Press.
- Durbin, J. and G. Watson (1950). "Testing for Serial Correlation in Least-Squares regression." *Biometrika* 37: 409-428.
- Edgerton, D. and G. Shukur (1999). "Testing Autocorrelation in a System Perspective." *Econometric Reviews* 18(4): 343-386.
- Fahrmeir, L. and H. Kaufman (1985). "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models." *The Annals of Statistics* 13(1): 342-368.
- Fujikoshi, Y. (1970). "Asymptotic Expansions of the Distributions of Test Statistics in Multivariate Analysis." *J. Sci. Hiroshima Univ. Ser. A-I* 34: 73-144.
- Fujikoshi, Y. (1988). "Comparison of Powers of a Class of Tests for Multivariate Linear Hypothesis and Independence." *Journal of Multivariate Analysis*. 26: 48-58.
- Godfrey, L. G. (1978). "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables." *Econometrica* 46(6): 1303-1310.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton.
- Harville, D. A. (1997). *Matrix Algebra From a Statisticians Perspective*, Springer Verlag.
- Heck, D. L. (1960). "Charts of Some Upper Percentage Points of the Distribution of the Largest Characteristic Root." *Annals of Mathematical Statistics* 31: 625-642.



- Hosking, J. R. M. (1980). "The Multivariate Portmanteau Statistic." *Journal of American Statistical Association* 75: 602-608.
- Khatri, C. G. (1972). "On the Exact Finite Series Distribution of the Smallest or the Largest Root of Matrices in Three Situations." *Journal of Multivariate Analysis*. 2: 201-207.
- Lee, Y. (1971). "Distribution of the canonical Correlations and Asymptotic Expansions for Distributions of Certain Independence Test Statistics." *The Annals of Mathematical Statistics* 2: 526-537.
- Ljung, G. M. and G. E. P. Box (1978). "On a measure of lack of fit in time series models." *Biometrika* 65: 297-303.
- Lutkepohl, H. (1993). Introduction to Multiple Time Series Analysis. Berlin, Springer-Verlag.
- Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. New York, Wiley.
- Perlman, M. D. and I. Olkin (1980). "Unbiasedness of Invariant Tests for MANOVA and Other Multivariate Problems." *The Annals of Statistics* 8(6): 1326-1341.
- Pillai, K. C. S. (1977). "Distributions of Characteristic Roots in Multivariate Analysis Part II. Non-null Distributions." *The Canadian Journal of Statistics* 5(1): 1-62.
- Rao, C. R. (1973). Linear Statistical Inference. New York, Wiley.
- Roy, S. N. (1945). "The Individual Sampling Distribution of the Maximum, the Minimum and any Intermediate Root of the Null Hypothesis." *Sankhya* 7: 133-158.
- Schmidt, P. (1976). Econometrics, Marcel Dekker.
- Shao, J. (1999). Mathematical Statistics, Springer.
- Shukur, G. (1997). Some Aspects of Statistical Inference in Systems of Equations, PhD Thesis, Dept of Statistics. Lund University, Sweden.
- Srivastava, V. K. and D. A. Giles (1987). Seemingly Unrelated Regression Equations Models, Marcel Dekker.



## Research Report

- |        |                                      |   |
|--------|--------------------------------------|---|
| 2002:1 | Frisén, M. & Sonesson, C.:           | Optimal surveillance. Based on exponentially weighted moving averages.                                    |
| 2002:2 | Frisén, M.:                          | Statistical surveillance. Optimality and methods.   |
| 2002:3 | Jonsson, R. & Persson, A.:           | Bayes prediction of binary outcomes based on correlated discrete predictors.                              |
| 2002:4 | Persson, A.:                         | Prediction of work resumption among men and women with lower back- and neck pain in a Swedish population. |
| 2002:5 | Persson, A.:                         | Prediction of work resumption in theory and practice.   |
| 2002:6 | Andersson, E., Bock, D. & Frisé, M.: | Detection of turning points in business cycles.   |
| 2002:7 | Andersson, E., Bock, D. & Frisé, M.: | Some statistical aspects on methods for detection of turning points in business cycles.                   |
| 2002:8 | Andersson, E., Bock, D. & Frisé, M.: | Statistical surveillance of cyclical processes with application to turns in business cycles.              |
| 2002:9 | Holgersson, T.:                      | Testing for non-normality in multivariate regression with nonspherical disturbances.                      |
| 2003:1 | Holgersson, T. & Shukur, G.:         | Testing for multivariate heteroscedasticity.  |