Research Report
Department of Statistics
Göteborg University
Sweden

Testing for non-normality in multivariate regression with nonspherical disturbances

Thomas Holgersson

Research Report 2002:9
ISSN 0349-8034

Mailing address:	Fax	Phone	Home Page:
Dept of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

Testing for Non-normality in Multivariate Regression with Nonspherical Disturbances

By H. E. T. Holgersson

Department of Statistics
Göteborg University
P.O. Box 660
SE-405 30 Göteborg, Sweden

Abstract: Statistical diagnostic testing is often associated with erratic conclusions due to the fact that a test against one certain specification may be highly sensitive to another specification. This paper concerns assessing normality of autocorrelated or heteroscedastic variables. It is shown why the type I error of skewness/kurtosis test limits 100% if the data are not i.i.d. We propose a set of tests for non-normality, which are robust to autocorrelation/heteroscedasticity, covering a wide class of situations. The size and power of the tests are investigated by Monte Carlo techniques.

Keywords: tests of non-normality, multivariate analysis, heteroscedasticity, autocorrelation.

I. Introduction

Statistical modelling often concerns diagnostic testing. Examples of such tests are tests for autocorrelation or heteroscedasticity. For instance, the analyst may ask "do I have reason to believe that my data are autocorrelated"? Certainly, a statistical technique capable of testing such a hypothesis is of great importance. But suppose that this test indicates autocorrelation only because another, (irrelevant) effect is present. Then the practitioner may be totally misled, trying to re-specify his/her model in a wrong direction.

Problems of this kind have been given surprisingly little attention, though some have been recognised. For example, tests for ARCH (auto regressive conditional heteroscedasticity) effects are well known to be sensitive to autocorrelation. Further, Horswell and Looney (1993) showed that tests based on skewness coefficients do not discriminate between skewed and (non-normal) symmetric distributions, and Koenker and Basset (1982) refers to the "shakily" fact that most tests for heteroscedasticity rest on the assumption of normality.

These effects of confounding suggest that research regarding diagnostic tests should involve robustness, rather than solely focus on size/power properties. Fortunately, some such results exist. For example, the robustness of the renowned Breusch-Godfrey (Breusch (1978), Godfrey (1978)) test for autocorrelation have been examined by Shukur (2000), who concludes that the test is asymptotically invariant to non-normality. Koenker and Basset (1982) suggest a non-normality roust heteroscedasticity test. These and other results on robustness demonstrating the possibility to (at least roughly) establish whiteness of data, regardless non-normality, are obviously of great interest. The opposite problem, i.e. to assess normality of non-i.i.d. data is more troublesome, as most non-normality tests are non-robust to autocorrelation. One specific example is given in Holgersson and Shukur (2001), where Monte Carlo simulations reveal that skewness/kurtosis tests for non-normality are highly sensible to autocorrelation, especially in large samples and high-dimensional variables, in the sense that the size of the non-normality test limits 100% when autocorrelation is present. Similarly, one may expect heteroscedasticity to

affect the properties of non-normality tests (which indeed it does, as we will see in Section III).

This paper concerns assessing normality of (possibly multivariate) regression models. We give an explicit reason why skewness/kurtosis tests are inconsistent in the presence of autocorrelation or heteroscedasticity (in the sense that the type-I error approaches 100% as the sample size increases). Furthermore, we propose three classes of tests for non-normality when heteroscedasticity or autocorrelation is present. The tests use different levels of information of the data generating process. The first assumes known covariance matrix, the second only assumes known structure of the covariance matrix while the third is useful for time series where the autocorrelation generating model is unknown. The size, power and robustness of the tests are then examined by means of Monte Carlo experiments.

The paper is organised as follows. In the next section we present the model we analyse. In Section III we show why standard tests for skewness and kurtosis, e.g. the Jarque and Bera (1987) test, will not converge to its asymptotic null distribution when the variable is non-i.i.d. Section IV is concerned with possible choices of observable proxy variables to the unobservable disturbances. In Section V we present a class of skewness/kurtosis tests which are variations of the Lütkepohl and Theilen (1991) test for multivariate non-normality on i.i.d. variables. In Section VI we present the design of the Monte Carlo experiment, while the results concerning size and power are presented in Section VII. Finally, a brief summary is given together with some conclusions in Section VIII.

II. Model specification

The model considered in this paper is the multivariate regression model

$$\mathbf{Y}_{(n\times G)} = \mathbf{X}_{(n\times k)}\mathbf{B}_{(k\times G)} + \boldsymbol{\varepsilon}_{(n\times G)}, \tag{2.1}$$

where n is the number of observations and G is the number of response variables. By assumption, $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2\boldsymbol{\Omega}_{(n\times n)}$ for some scalar $\sigma^2 < \infty$, $\boldsymbol{\Omega}$ a positive definite (p.d.) matrix, and $E[\mathbf{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}] = \mathbf{0}_{(k\times G)}$. Furthermore, define $\tilde{\mathbf{X}} := \boldsymbol{\Omega}^{-1/2}\mathbf{X}$. We then assume that $plim(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}/n)$ is a finite and non-singular matrix. Some further (not very restrictive) assumptions will be added later on.

The model of (2.1) is frequently used, for example in the context of allocation models (see e.g. Bewley (1986)), and is a generalisation of univariate multiple regression (i.e. multivariate regression with only one equation) and is hence quite general.

A frequently made assumption in (2.1) is that $\boldsymbol{\varepsilon}$ is normally distributed. That a variable is normally distributed is a somewhat vague property. This may lead the practitioner to believe that a non-normality test is assessing whether a variable belongs to the family of normal distributions or not. This is usually not the case as most standard tests are, under the null hypothesis, assuming that $\boldsymbol{\varepsilon}$ is *i.i.d.* normal. Thus we use two assumptions in the null hypothesis, which is often overlooked.

The density function of a multivariate normally distributed variable is given by

$$f(\boldsymbol{\varepsilon}') = \left[(2\pi)^2|\boldsymbol{\Sigma}|\right]^{-\left(\frac{n}{2}\right)}|\boldsymbol{\Omega}|^{-\left(\frac{G}{2}\right)} exp\left[tr\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}\right)\right], \text{ or } \boldsymbol{\varepsilon}' \sim N_{G,n}\left(\mathbf{0},\ \boldsymbol{\Sigma}_{(G\times G)}, \boldsymbol{\Omega}_{(n\times n)}\right).$$

(see Srivastava (1979)). The covariances between the G equations is determined by $\boldsymbol{\Sigma}_{(G\times G)}$ and the covariance matrix of the n observations is $\boldsymbol{\Omega}_{(n\times n)}$. Whenever $\boldsymbol{\Omega} \neq \sigma^2\mathbf{I}$ for scalar σ^2, the errors are said to be *nonspherical.* Two such cases that will be

considered in detail are heteroscedasticity and autocorrelation. Autocorrelation is usually associated with data sampled over time-equidistant intervals, while heteroscedasticity often arises in a model due to the fact that the variance is a function of the regressors.

III. Failure of skewness and kurtosis testing of non-i.i.d. variables

The skewness and kurtosis coefficients are frequently used to test for non-normality. Particularly popular is the Jarque and Bera (1987) omnibus test (JB), with its statistic defined as $\{n\hat{\gamma}^2{}_1/6\}+\{n(\hat{\gamma}_2-3)^2/24\}$, where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the sample skewness and kurtosis coefficients. When the target variable is i.i.d. normally distributed, the statistics above are each asymptotically $\chi^2_{(1)}$ distributed. If the observations are not independent, i.e. if $\mathbf{\Omega}\neq\mathbf{I}$ due to autocorrelation, the JB statistic will not converge to its null distribution, as is exemplified in Holgersson and Shukur (2001). To obtain an explicit reason for the non-convergence of the statistic in this case, we use the result of Lomnicki (1961): Consider a univariate time series $x_i=\sum_{k=0}^{\infty}\psi_k\varepsilon_{i-k}$, where $\sum_{k=1}^{\infty}|\psi_k|<\infty$ and $\varepsilon_i\sim iid\ N(0,\ \sigma^2)$ with autocorrelations $\rho_k:=E(x_ix_{i-k})/E(x_i^2)$. Lomnicki showed that, for a one-dimensional random variable, $\{\sqrt{n}\,\hat{\gamma}_1/\sqrt{6}\}\xrightarrow{\ell}N\left(0,\sum_{k=-\infty}^{k=\infty}\rho_k^3\right)$ and $\{\sqrt{n}(\hat{\gamma}_2-3)/\sqrt{24}\}\xrightarrow{\ell}N\left(0,\sum_{k=-\infty}^{k=\infty}\rho_k^4\right)$, where $\xrightarrow{\ell}$ denotes convergence in law. From this follows immediately that

$$\rho_k\neq 0\Rightarrow\{n\hat{\gamma}_1/6\}+\{n(\hat{\gamma}_2-3)^2/24\}\not\xrightarrow{\ell}\chi^2_{(2)}\quad\forall_{k\neq 0}. \tag{3.1}$$

For example, consider a process defined by $x_i=\theta x_{i-1}+\varepsilon_i$, i.e. the classical MA(1) process. Then it is well known that $\rho_1=-\theta/(1+\theta^2)$, $\rho_k=0$, $k>1$. Hence $\hat{\gamma}_1\xrightarrow{\ell}N\left(0,\left\{-\theta^3/(1+\theta^2)^3\right\}\right)$ and the non-convergence follows directly for $\theta\neq 0$. To examine the null distribution of the JB statistic for the case where $\mathbf{\Omega}\neq\mathbf{I}$ due to heteroscedasticity rather than autocorrelation, we use the result of Kendall and Stuart (1976), that, given normality,

$$E\left[\left(\frac{m_3}{m_2^{3/2}}\right)^d\right]=\frac{E\left[m_3^d\right]}{E\left[m_2^{3d/2}\right]}, \qquad \text{and} \qquad E\left[\left(\frac{m_4}{m_2^2}\right)^d\right]=\frac{E\left[m_4^d\right]}{E\left[m_2^{d2}\right]},$$

where $m_j=(1/n)\sum_{i=1}^{n}(x_i-\bar{x})^j$ and $\bar{x}$ is the sample mean. The moments of the univariate skewness and kurtosis coefficients for heterogeneous observations can then be obtained by assuming that the random variables $\{x_i\}_{i=1}^{n}$ are independently distributed $N(\mu,\sigma_i^2)$. The numerators and denominators of the quotients above are derived in Appendix A (using leading term approximations):

$$E\left[m_3^2\right]=\frac{1}{n^2}\left[15\sum_{i=1}^{n}\sigma_i^6+(9/n^2)\left(\sum_{i\neq j\neq k}\sigma_i^2\sigma_j^2\sigma_k^2\right)-(18/n)\left(\sum_{i\neq j}\sigma_i^4\sigma_j^2\right)\right]+o\left(n^{-1}\right). \tag{3.2}$$

$$E\left[m_4\right]=3\sum_{i=1}^{n}\sigma_i^4/n+o\left(n^{-1/2}\right). \tag{3.3}$$

$$E\left[m_2^2\right]=\sum_{i\neq j}\sigma_i^2\sigma_j^2/n^2+o\left(n^{-1/2}\right). \tag{3.4}$$

$$E\left[m_2^3\right]=\sum_{i\neq j\neq k}\sigma_i^2\sigma_j^2\sigma_k^2/n^3+o\left(n^{-1/2}\right). \tag{3.5}$$

From these expressions it follows that if $\sigma_1^2\neq\sigma_2^2\neq\ldots\neq\sigma_n^2$ then the variance of $\sqrt{n}\hat{\gamma}_1$, i.e. n(3.2)/(3.5), will in general not limit the value 6, and the expectation of $\sqrt{n}(\hat{\gamma}_2-3)$, i.e. $\sqrt{n}\left((3.3)/(3.4)-3\right)$, will not limit the value 0. Hence,

$$\left(\sigma_1^2\neq\sigma_2^2\neq\ldots\neq\sigma_n^2\right)\Rightarrow\left\{n\hat{\gamma}_1/6\right\}+\left\{n\left(\hat{\gamma}_2-3\right)^2/24\right\}\overset{\ell}{\not\to}\chi^2_{(2)}. \tag{3.6}$$

For example, consider a random variable such that $x_i\sim N(\mu,\sigma_i^2)$ where $\sigma_i^2=i/n$ for $i=1,\ldots,n$. Then it may be shown that $\lim_{n\to\infty}\sqrt{n}(\hat{\gamma}_2-3)=\infty$ (Appendix A3).

In other words, measures of multivariate skewness and kurtosis that are extensions of those above will in general not converge to their null distribution when the covariance of the disturbance vector is non-scalar (e.g. Lütkepohl and Theilen (1991), Malkovich

and Afifi (1973) or Jarque and Mckenzie (1995)). Indeed, it is likely that most standard non-normality tests, for example those based on empirical distribution functions or empirical characteristic functions, will not converge to their null distributions if the data is not i.i.d., as they usually measure the complete distribution of the target variable. Consequently, accepted null hypothesis will indicate that the data is normally distributed, while a rejected null hypothesis may either be due to non-normality *or* non-scalar covariance matrix $\boldsymbol{\Omega}$. Thus, standard tests for non-normality are diagnostic only if we know that the target variable is i.i.d. Since non-scalar covariance matrixes are frequently occurring in various applications of statistics, this confounding problem is rather serious. The important issue is then to overcome the problem. There are two obvious possibilities. One is to construct tests that are robust to non-i.i.d. data (i.e. to construct tests that assess the distribution of $\boldsymbol{\varepsilon}$). The second is to assess the distribution of transformed variables with scalar covariance matrices (i.e. to asses the distribution of $\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}$). We shall consider both approaches in this paper. Details are given in Section V.

IV. Proxy variables to the unobservable disturbances

Often we want to make some inferences of the disturbance component $\boldsymbol{\varepsilon}$ of (2.1). As this is unobservable we will have to use an observable proxy variable instead, usually the residuals $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}$. A specifically useful property of residuals that ensures their usefulness in diagnostic testing is that they converge in probability to the true disturbances, i.e. $\left|\hat{\varepsilon}_{i,n} - \varepsilon_i\right| \xrightarrow{p} 0$. Lütkepohl and Schneider (1989) examined the possibility of using residuals for evaluation of the distribution of the random component for the case of pure autoregressive processes. They concluded that the residual-based test works very well. In our case with exogenous information, a natural approach is to use GLS/FGLS estimation, thus covering heteroscedasticity as well.

Apart from difficulty with assessing normality, the main problem with non-spherical disturbances is that the point estimates of regression parameters will be inefficient and the interval estimates may be erratic. In such situations the GLS method is commonly

used. For multivariate regression we can describe the method as follows. Let $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2\boldsymbol{\Omega}$. If $\boldsymbol{\Omega}$ is known, we can perform the transformation

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon} \mapsto \boldsymbol{\Omega}^{-1/2}\mathbf{Y} = \boldsymbol{\Omega}^{-1/2}\mathbf{X} + \boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}, \text{ or, } \tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\mathbf{B} + \tilde{\boldsymbol{\varepsilon}}. \tag{4.1}$$

The OLS estimates of the regression parameters of the transformed model above are well known to be both unbiased and efficient. The residuals of (4.1) are usually specified as $\mathbf{Y} - \hat{\tilde{\mathbf{Y}}}$ where $\hat{\tilde{\mathbf{Y}}} = \tilde{\mathbf{X}}\hat{\mathbf{B}}_{OLS} = \tilde{\mathbf{X}}\left\{\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}\tilde{\mathbf{Y}}\right\}$. These residuals will limit $\boldsymbol{\varepsilon}$ in distribution, which are non-spherical. Therefore, we will instead use the residuals $\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS} := \tilde{\mathbf{Y}} - \hat{\tilde{\mathbf{Y}}}$. It is shown in Appendix B that $\hat{\tilde{\boldsymbol{\varepsilon}}}_i \xrightarrow{\ell} \tilde{\boldsymbol{\varepsilon}}_i$. These residuals are particularly useful in testing for non-normality, since the covariance matrix of $\tilde{\boldsymbol{\varepsilon}}$ is scalar (i.e. $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2\mathbf{I}$). The transformation matrix $\boldsymbol{\Omega}^{-1/2}$ for heteroscedastic variables may for example consist of known weights due to unequal number of observations in each point $\mathbf{Y}_i$ (e.g when $\mathbf{Y}_i$ is an average of n_i individuals, say). The GLS transformation matrix for a regression model with AR(p) disturbances is well known and can be written as $\boldsymbol{\Omega}^{-1/2} = h(\boldsymbol{\phi})$, where $h(\cdot)$ is a function given by the autoregression parameters ϕ. Gailbraith and Zinde-Walsh (1992) supply exact and asymptotic expressions for $h(\cdot)$ for a general stationary ARMA process. Now, the residuals $\hat{\tilde{\boldsymbol{\varepsilon}}}$ mentioned above are not always realistic, particularly not for the case of autocorrelation, since $\boldsymbol{\Omega}$ is often unknown. For example, the heteroscedasticity is sometimes assumed to be of some parametric form, i.e. $\omega_{ii,g} = \gamma E[Y_{ig}]$ as in Bickel (1978), and has to be estimated. Denoting the resulting estimated covariance matrix by $\hat{\boldsymbol{\Omega}}$, we may perform the same transformation as in (4.1),

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon} \mapsto \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{Y} = \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{X} + \hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{\varepsilon}, \text{ or, } \hat{\mathbf{Y}} = \hat{\mathbf{X}}\mathbf{B} + \hat{\boldsymbol{\varepsilon}}, \text{ say.} \tag{4.2}$$

For multivariate FGLS to operate in the case of autocorrelation, we assume that the parameters are equal in all processes. For example, if the disturbances follows an AR(1) process with parameter ϕ, then by assumption $\phi_1 = \ldots = \phi_G$. We may then take

the mean value of the estimates above as our final estimate: $\bar{\hat{\phi}} = \sum_{g=1}^{G} \hat{\phi}_g / G$. The residuals of (4.2) are defined as $\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS} := \hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{OLS}$. It may be shown that $\hat{\tilde{\boldsymbol{\varepsilon}}}_i \xrightarrow{\ell} \tilde{\boldsymbol{\varepsilon}}_i$, (see Appendix C). Hence, if the assumptions regarding $\boldsymbol{\Omega}$ and $\hat{\boldsymbol{\Omega}}$ of (4.1) and (4.2) above hold, the $\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS}$ and $\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}$ residuals will be useful for testing non-normality. The situation where no explicit information of $\boldsymbol{\Omega}$ is available is more difficult. To our knowledge, no techniques are at the present available for assessing non-normality of heteroscedastic data of unknown structure. However, the case of autocorrelation can be handled by using some quite general assumptions of the process, even if the autocorrelation generating model is unknown. Our three classes of tests are presented and discussed in the next section.

V. Assessing normality of non-spherical variables

Skewness and kurtosis tests are popular for assessing normality. The reasons for this may be that they are easy to implement and are often included in program packages. We briefly mentioned some multivariate measures in Section III. Here we will consider variations of a test suggested by Lütkepohl and Theilen (1991) that we feel is particularly useful for non-spherical disturbances. The test is a multivariate generalisation of measures of univariate Skewness and Kurtosis, defined as follows: Let $\mathbf{U} = [\mathbf{u}_1 ... \mathbf{u}_n]$ be a sample of n observations on a G variate random vector, and let $\bar{\mathbf{u}}$ and $\mathbf{S}$ be the corresponding sample mean and covariance matrix respectively. Further, let $\mathbf{P}$ be the Cholesky decomposition of $\mathbf{S}$ (such that $\mathbf{S} = \mathbf{PP}'$), and define

$$\mathbf{w}_i := [w_{1i} ... w_{Gi}]' = \mathbf{P}^{-1}(\mathbf{u}_i - \bar{\mathbf{u}}),$$

$$b_1 := [b_{11} ... b_{G1}]' \text{ with } b_{g1} = \sum_{i=1}^{n} w_{ig}^3 / n,$$

$$b_2 := [b_{12} ... b_{G2}]' \text{ with } b_{g2} = \sum_{i=1}^{n} w_{ig}^4 / n,$$

for $g=1,...,G$. Then, under the null hypothesis of normality, we have $\tilde{b}_1 := \left(nb_1'b_1/6\right) \xrightarrow{\ell} \chi^2_{(G)}$ and $\tilde{b}_2 := \left(n(b_2-3)'(b_2-3)/24\right) \xrightarrow{\ell} \chi^2_{(G)}$ and it follows that

$$\tilde{b} := \tilde{b}_1 + \tilde{b}_2 \xrightarrow{\ell} \chi^2_{(2G)} \tag{5.1}$$

(see Lütkepohl and Theilen (1991)). We will refer to this test as the LT test. It has been shown that the limiting distribution of (5.1) also follows when $\mathbf{u}_i$ is replaced by residuals from an VAR(p) process (Lütkepohl and Theilen (1991)), and as the VAR(p) is a more general model than the multivariate regression model, their proof covers the multivariate regression as well. In order to distinguish between different target variables of the statistic we shall henceforth write (5.1) as $\tilde{b}(\mathbf{U})$.

The null hypothesis H_0: $\mathbf{U} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ is rejected at the α-level whenever

$$\tilde{b}(\mathbf{U}) > \eta \text{ where } P\left(\tilde{b}(\mathbf{U}) > \eta \mid \mathbf{U} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})\right) = \alpha .$$

It is readily seen that this test will have power against any distribution such that $E\left[b_{g1}\right] \neq 0$ *or* $E\left[b_{g2}\right] \neq 3$ *or* $V\left[b_{g1}\right] \neq 6$ *or* $V\left[b_{g2}\right] \neq 24$ if the $\mathbf{u}_i$*'s* are i.i.d. Note though, that according to (3.1) and (3.6), we need $\boldsymbol{\Omega} = \mathbf{I}$ for (5.1) to hold. As the limiting null distribution of our residuals $\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}$ and $\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS}$ in Section **IV** equals that of the true disturbances, though with scalar covariance matrix, both $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)$ and $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS}\right)$ can be used to test for non-normality. However, it is shown in Holgersson and Shukur (2001) that the size of residual-based tests might be biased in small samples, due to the distributional properties of **X**. In addition, the LT statistic itself has unknown null distribution and relies on asymptotic properties. However, for the case of known $\boldsymbol{\Omega}$, we can actually obtain an exact test. As the statistic of (5.1) is invariant to linear transformations (see Lütkepohl and Theilen (1991)), the size may be controlled by Monte Carlo techniques in the following way: Let $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0$ be the

fixed realised test statistic defined above, and suppose that we have B independent realisations of $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0$ at hand, $\left\{b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_6\right\}_{6=1}^{B}$ say, and let

$$\hat{p}_B\left(b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right) := \frac{\#\left\{b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_6 \geq b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right\}+1}{B+1}.$$

The associated (Monte Carlo) critical region is then defined as $\hat{p}_B\left(b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right) \leq \alpha$, so that $\hat{p}_B\left(b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right)$ may be interpreted as an estimated p-value of the realised statistic $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0$. If $\alpha(B+1)$ is an integer, we have $P\left[\hat{p}_B\left(b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right) \leq \alpha\right] = \alpha$ (see Dufour, Abdeljelil et al. (1998)). The rationale of this technique is that $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_6$ have the same distribution as (the population counterpart of) $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0$. Now, since $\boldsymbol{\varepsilon}$ is unobservable, we can instead use the observable linear combination $\hat{\tilde{\boldsymbol{\varepsilon}}} = \left(\mathbf{I} - \tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\right)\tilde{\boldsymbol{\varepsilon}} =: \mathbf{M}_{\tilde{\mathbf{X}}}\tilde{\boldsymbol{\varepsilon}}$, i.e. our GLS residuals, to obtain a Monte Carlo critical region for $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)$ by noting that $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0 = b\left(\mathbf{M}_{\tilde{\mathbf{X}}}\tilde{\boldsymbol{\varepsilon}}\right)$ and taking $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_6 := b\left(\mathbf{M}_{\tilde{\mathbf{X}}}\boldsymbol{\delta}_6\right)$, where $\boldsymbol{\delta}_6$ is an $N_{G,n}\left(\mathbf{0}, \mathbf{I}_{(G\times G)}, \mathbf{I}_{(n\times n)}\right)$ Monte Carlo sample. Since $\boldsymbol{\delta}$ has the same distribution as $\tilde{\boldsymbol{\varepsilon}}$ under the null hypothesis, and $\mathbf{M}_{\tilde{\mathbf{X}}}$ is a fixed, known matrix, it follows that $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_6$ will have the same distribution as $b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)$ (see Appendix D). The null hypothesis H_0: $\boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}\right)$ is then rejected at the α-level whenever $\hat{p}_B\left(b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right) \leq \alpha$, i.e.

$$P\left(\hat{p}_B\left(b\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS}\right)_0\right) \leq \alpha \,|\, \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}\right)\right) = \alpha. \tag{5.2}$$

By using this approach for the GLS residuals we obtain an exact inference procedure that is equivalent to using the true unobservable disturbances in conjunction with exact critical values for the statistic. We will refer to this test as the *Monte Carlo* test.

For the case of only known *structure* of the covariance matrix $\mathbf{\Omega}$, the Monte Carlo method will no longer be exact as the $\mathbf{M}_{\tilde{\mathbf{X}}}$-part of $\hat{\tilde{\boldsymbol{\varepsilon}}} := \mathbf{M}_{\tilde{\mathbf{X}}}\tilde{\boldsymbol{\varepsilon}}$ will be stochastic in finite samples and would so result in a stochastic critical region. Therefore we will instead use $\tilde{b}\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS}\right)$ as a test statistic with its critical value (under the null hypothesis of normality) determined by η where $P\left(\chi^2_{(2G)} \geq \eta\right) = \alpha$. Hence,

$$\lim_{n\to\infty} P\left(\tilde{b}\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS}\right) > \eta \,|\, \boldsymbol{\varepsilon} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega}\right)\right) = \alpha . \tag{5.3}$$

This test will be referred to as the *Feasible Generalised Least Square Residual* test.

Finally, we will formulate a non-normality test in presence of autocorrelation of unknown type. From the expressions on p.5 we have

$$\left\{n\hat{\gamma}_1^2 \Big/ 6\sum_{k=-\infty}^{k=\infty}\rho_k^3\right\} \xrightarrow{\ell} \chi^2_{(1)} \text{ and } \left\{n\left(\hat{\gamma}_2 - 3\right)^2 / 24\sum_{k=-\infty}^{k=\infty}\rho_k^4\right\} \xrightarrow{\ell} \chi^2_{(1)}. \tag{5.4}$$

Proceeding just like in (5.1), though with

$$\tilde{b}_{1,\rho} := \left[\tilde{b}_{11} \ldots \tilde{b}_{G1}\right]' \text{ where } \tilde{b}_{g1} = \sum_{i=1}^{n} w_{ig}^3 \Big/ n\sqrt{\sum_{k=-\infty}^{k=\infty}\rho_{k,g}^3}\,, \text{ and} \tag{5.5}$$

$$\tilde{b}_{2,\rho} := \left[\tilde{b}_{12} \ldots \tilde{b}_{G2}\right]' \text{ where } \tilde{b}_{g2} = \sum_{i=1}^{n} w_{ig}^4 \Big/ n\sqrt{\sum_{k=-\infty}^{k=\infty}\rho_{k,g}^4}\,. \tag{5.6}$$

we obtain $\tilde{b}_\rho := \tilde{b}_{1,\rho} + \tilde{b}_{2,\rho} \xrightarrow{\ell} \chi^2_{(2G)}$. Now, the ρ_k's are unknown and have to be replaced by consistent estimates. Consistent estimates of $\sum_{k=-\infty}^{k=\infty}\rho_{k,g}^3$ and $\sum_{k=-\infty}^{k=\infty}\rho_{k,g}^4$ cannot be achieved for all kind of processes. However, consistent estimates for a wide class of processes may be obtained by replacing all ρ_k, $k = 1, 2, \ldots, q$, with its least square estimates (e.g. Hamilton (1994)) and putting ρ_k $(k > q)$ equal to zero. The truncation point q cannot be chosen arbitrarily large, but it may be allowed to grow with n as long as q grows at a slower rate than $\sqrt{n}$. For example, consistent

estimates of $\sum_{k=-\infty}^{k=\infty}\rho_{k,g}^3$ and $\sum_{k=-\infty}^{k=\infty}\rho_{k,g}^4$ are obtained for $q=n^{4/10}$ (see Appendix E), though possible a-priori knowledge of the autocorrelation structure should be used to set a fixed truncation point if possible. For example, most of the "regular" processes used in the literature, such as e.g. stationary ARMA processes, have autocorrelations close to zero already after 10-20 lags. Our test statistic for non-normality for data with unknown autocorrelation will be $\tilde{b}_{\hat{\rho}} := \tilde{b}_{1,\hat{\rho}} + \tilde{b}_{2,\hat{\rho}} \overset{\ell}{\rightarrow} \chi^2_{(2G)}$, i.e.

$$\lim_{n\to\infty} P\left(\tilde{b}_{\hat{\rho}}\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{OLS}\right) > \alpha \,|\, \boldsymbol{\varepsilon} \sim N\left(\mathbf{0},\boldsymbol{\Sigma},\boldsymbol{\Omega}\right)\right) = \alpha\,. \tag{5.7}$$

This test will be referred to as the *Feasible Lomniki* test.

VI. The Monte Carlo design

In this section we discuss some characteristics that are involved in the problem of testing for non-normality in multivariate regression with non-spherical disturbances of known and unknown structures as in the situations mentioned in Section **V**. The small sample properties of our tests treated in Section **V** are unknown. It is therefore important to examine whether the actual behaviour of these tests is adequately approximated by asymptotic theory. In the absence of exact results, it is necessary to investigate the finite sample performance of the tests by means of simulation experiments. When investigating the properties of a classical test procedure, two aspects are of prime importance. Firstly, we wish to see if the actual size of the test is close to the nominal size (used to decide the critical region for the rejection of the null hypothesis). Given that the actual size is a reasonable approximation to the nominal size, we then wish to investigate the power of the test. In general, when comparing different tests we will therefore prefer those whose (i) actual size lies close to the nominal size and, given that (i) hold, (ii) have the greatest power. Other relevant criteria, such as which test has the most soundly theoretical basis, or which test is the most simple to perform, cannot be judged quantitatively. Therefore we leave this aspect to be judged by the reader.

In a Monte Carlo study we calculate the estimated size by observing how many times the null is rejected in repeated samples under conditions where the null is true. However, this estimated size is associated with a source of uncertainty due to a finite number of replications. Therefore we need a tool for deciding how close the estimated size must be to the nominal size for the result to be judged as "reasonable". It makes sense to define a test with nominal size 5% as reasonable if the estimated size lies between 4% and 6%. We then need to form a $\pm 1\%$ interval as a function of the number of replicates. An approximate confidence interval for the actual size (α) can be given by

$$\hat{\alpha} \pm 2\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{R}}$$

where $\hat{\alpha}$ is the estimated size and R is the number of replications. Solving for R for $\pm 1\%$ interval gives $R = 2250$. Let us say we choose $R = 10000$. This gives us a conservative interval of $\pm 1\%$. Hence, if the estimated size of any of our tests exceeds the interval $0.06 - 0.04$, for $\alpha = 0.05$, $R = 10000$, we conclude that the true size of the test systematically exceeds the nominal size. Note though that this interval is valid only when judging the estimated size for one single sample size. Judgement of the estimated size over a range of sample sizes simultaneously would require multiple inference techniques.

Given that the actual size is a reasonable approximation of the nominal size, we then wish to investigate the actual power of the test (i.e. the probability of rejecting the null when false). We will therefore consider one skew and one symmetric alternative distribution. The first is defined by $\mathbf{X} = \mathbf{LY}$ where $\mathbf{Y}_i = [Y_{1i},...,Y_{Gi}]'$ $i = 1,...,n$ are i.i.d. $\chi^2_{(10)}$ variates, and $\mathbf{LL}' = \mathbf{\Sigma}_{G\times G}$, a p. d. covariance matrix. We will write this distribution as $\chi^2_{(10,\Sigma)}$. In order to obtain heavy-tailed disturbances we use a variable defined by $\mathbf{X} = \mathbf{LT}$, where $\mathbf{T}_i = [T_{1i},...,T_{Gi}]'$ are i.i.d. student $T_{(5)}$ distributed variates, and again $\mathbf{LL}' = \mathbf{\Sigma}$. We write this distribution as $T_{(5,\Sigma)}$.

Further, in Holgersson and Shukur (2001) the simulation results indicate that the biases in the estimated skewness and kurtosis, due to using residuals as a proxy to $\boldsymbol{\varepsilon}$, are negligible when $\mathbf{X}$ is close to a normal distribution. However, when the regressors

are following a heavy-tailed distribution, the rate of convergence to its expectations is rather slow. We will therefore use the $T_{(3,\Sigma)}$ distribution in the regressors in order to obtain more general regressors.

In order to examine the effects of autocorrelated noise, we need to choose some forms of autocovariances. For simplicity, we consider (marginal) AR(1) processes, i.e. $\varepsilon_{ig} = \phi\varepsilon_{i-1,g} + \delta_{ig}$, which have a fairly long "memory", the autocorrelations being $\rho_k = \phi^k$. The heteroscedastic disturbances will be examined using covariance matrices that are functions of a subset of the regressors, namely $\sigma_i^2 = \sigma^2 + v_1 X_{1i} + v_2 X_{2i}$. Finally, preliminary examinations of the Lütkepohl and Theilen (1991) statistic of (5.1) revealed a rather slow convergence to its asymptotic null distribution, resulting in unnecessarily poor small sample properties. Hence we have used the invariance of linear transformations of the statistic to simulate empirical critical values, using one million replicates. The empirical critical values along with a graph showing the true size of the test are given in Appendix F. All tests examined in the Monte Carlo experiment are based on the formulas (5.2), (5.3) and (5.7) with the 95% quantile replaced by the simulated (nearly) exact 95% quantile of the statistic. Below we present a table containing the relevant factors used in the experiment.

TABLE A. The tests used in the experiment.

Heteroscedasticity robust tests	Autocorrelation robust tests
Monte Carlo (MC)	Monte Carlo (MC)
Feasible Generalised Lest Square Residuals (FGLSR)	Feasible Generalised Lest Square Residuals (FGLSR)
	Feasible Lomninki test (FL)

TABLE B. Properties of factors in the experiment.

Factor	Symbol	Value/property
Numbers of regressors	k	5
Distribution of regressors	$\mathbf{X}_{[1:5]}$ $\mathbf{X}_{[6:7]}$	$T_{(3)}$ * U(1, 10)
Nominal size	α	5%
Number of equations	G	5
Sample size	n	15, 35, 65, 100, 150, 200, 500.
Distribution of disturbances	$\boldsymbol{\varepsilon} \sim$	$N(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon}, \boldsymbol{\Omega})$, $T_{(5,\Sigma_{\varepsilon})}$, $\chi^2_{(10,\Sigma_{\varepsilon})}$ *
Structure of heteroscedasticity	$\boldsymbol{\Omega}$	$diag\left(\mathbf{X}_{[6:7]}\boldsymbol{\theta}_{(2\times1)}\right)$
Value of parameter in the AR(1) processes	ϕ	0.2 0.7
Value of heteroscedasticity parameters	$\boldsymbol{\theta}_{(2\times1)}$	$[3\ 2]'$ $[1\ 0]'$
Number of resamples in the Monte Carlo test.	B	199

* The toeplitz operator provides a convenient technique for choosing the elements in high-dimensional matrix-valued parameters. It is a mapping $\mathbb{R}^G \mapsto \mathbb{R}^{G^2}$ where the upper triangle of the symmetric matrix $\mathbf{A} = toeplitz(\psi_1, ..., \psi_G)$ is defined by its diagonal elements $\mathbf{A}_{gg} = \psi_1$, $g = 1, ..., G$, the first off-diagonal by $\mathbf{A}_{g(g+1)} = \psi_2$, $g = 1, ..., (G-1)$, the second off-diagonal by $\mathbf{A}_{g(g+2)} = \psi_3$, $g = 1, ..., (G-2)$ and so on. Here, the covariance matrix of the regressors is $\boldsymbol{\Sigma}_x = toeplitz(5\ 3\ 1.8\ 0\ 0)$ and the covariance matrix of the disturbances are set to $\boldsymbol{\Sigma}_{\varepsilon} = toepliz(25\ 15\ 7\ 1.2\ 0.2)$.

VII. Results

In this section, we present our results of the main dominating effects in our Monte Carlo experiment regarding size and power properties of the variants of the LT test. The results will be presented in tabular form in two parts, *size properties* and *power properties* respectively. All simulations have been performed using R = 10 000 replicates.

Size properties

In this subsection we present our results concerning the size properties of the tests proposed in Section **V**. Table I shows the size properties of the test when the variables are i.i.d.. All tests perform fairly well except for the case of only 15 observations (a sample size which leaves only a tiny number of degrees of freedom). Expectedly, the Monte Carlo technique gives an exact size, with simulation fluctuations around the nominal size. Since $\mathbf{\Omega} = \mathbf{I}$ is a special case of a known covariance matrix, this Monte Carlo technique is more useful than it may seem, as scalar-valued covariance matrixes do exist in some situations (indeed, this is the situation examined in most empirical studies of non-normality tests).

Table II presents the size properties for our autocorrelation robust tests. As expected, the exact Monte Carlo technique yields correct size for all sample sizes. The most important property though, is that all three tests reach its nominal size asymptotically.

Moving on to the heteroscedasticity robust tests in Table III, we see that all tests have correct size with the exception of small deviations in the case of $n = 15$. What is particularly worth noting is that the "magnitude" of heteroscedasticity does not seem to matter.

Table I.

Size properties for i.i.d. disturbances					
	Autocorrelation robust tests			Heteroscedasticity robust tests	
n	**MC**	**FGLSR**	**FL**	**MC**	**FGLSR**
15	0,045	0,037	0,031	0,051	0,039
35	0,050	0,047	0,050	0,045	0,053
65	0,052	0,045	0,053	0,051	0,049
100	0,051	0,052	0,049	0,052	0,055
150	0,048	0,050	0,050	0,048	0,051
200	0,049	0,048	0,050	0,053	0,050
500	0,050	0,047	0,051	0,052	0,053

Table II.

Size properties for autocorrelated disturbances						
	$\rho = 0.2$			$\rho = 0.7$		
n	**MC**	**FGLSR**	**FL**	**MC**	**FGLSR**	**FL**
15	0,048	0,045	0,030	0,056	0,046	0,028
35	0,051	0,047	0,044	0,050	0,048	0,021
65	0,048	0,049	0,048	0,052	0,052	0,022
100	0,049	0,047	0,050	0,051	0,050	0,028
150	0,048	0,055	0,049	0,050	0,051	0,031
200	0,052	0,053	0,048	0,051	0,050	0,036
500	0,051	0,051	0,052	0,051	0,051	0,048

Table III.

Size properties for heteroscedastic disturbances				
	$\sigma_i^2 = X_{6,i}$		$\sigma_i^2 = 3X_{6,i} + 2X_{7,i}$	
n	**MC**	**FGLSR**	**MC**	**FGLSR**
15	0,054	0,035	0,050	0,034
35	0,052	0,051	0,051	0,050
65	0,050	0,054	0,055	0,049
100	0,048	0,057	0,048	0,048
150	0,049	0,056	0,051	0,052
200	0,051	0,058	0,050	0,050
500	0,048	0,052	0,049	0,052

Power properties

In this section we present our results regarding the power properties of the various tests of Section **V**. Tables IV-VI concerns the power properties of our autocorrelation robust and heteroscedasticity robust tests for the $\chi^2_{(10,\Sigma)}$ distribution. Table IV examines the case when the data are i.i.d.. All three autocorrelation robust tests behave similar. A particularly interesting finding is that the FL test has the same power as the exact MC test. Moving on to Table V we see that the power becomes slightly lower for low autocorrelation ($\phi = 0.2$) as compared to the i.i.d. case in Table IV. Increasing the autocorrelation to $\phi = 0.7$ causes further power reduction. In particular, the FL test needs 500 observations to reach full power as compared to the i.i.d. case when full power was reached for $n = 100$. What is striking though, is that the level of heteroscedasticity does not seem to matter. The tests behave similar for low and high heteroscedasticity, according to Table VI. Also, note that the power of the FGLSR and the MC tests are almost identical. In Tables VII-IX we examine the power properties of our tests for the $T_{(5,\Sigma)}$ distribution using the same autocorrelation and heteroscedasticity settings as in Tables IV-VI. In Table VII we see that the three autocorrelation robust tests have almost identical behaviour over the different sample sizes, and so does the two heteroscedasticity robust tests. The powers reported in Table VIII are similar to that of Table V. The power properties do not differ among the tests for low autocorrelation while for high autocorrelation the FL test clearly has lower power than the MC and the FGLSR test. Further, in Table IX, the difference between the FGLSR and the MC test seems to be minimal, and the intensity of heteroscedasticity, again, has no impact on the tests.

To sum up, the power properties behave quite expectedly. All three autocorrelation robust tests consistently detect the non-normal distributions. In general, the factors that affect the power properties of our tests proved to be similar to those that affect the size. For high autocorrelation though, there is a clear distinction between the tests in the sense that the test using most information of the data generating process (the MC test) has highest power, while the test using no information of the data generating process (the FL test) has lowest power. The same result does not hold for the heteroscedasticity robust tests, as the two tests appear to have the same power.

Table IV.

Power properties for i.i.d. disturbances, $\delta \sim \chi^2_{(10,\Sigma)}$					
	Autocorrelation robust tests			Heteroscedasticity robust tests	
n	**MC**	**FGLSR**	**FL**	**MC**	**FGLSR**
15	0,041	0,034	0,027	0,045	0,029
35	0,434	0,435	0,426	0,440	0,428
65	0,938	0,945	0,947	0,934	0,936
100	1	1	1	1	0,998
150	1	1	1	1	1
200	1	1	1	1	1
500	1	1	1	1	1

Table V.

Power properties for autocorrelated disturbances, $\delta \sim \chi^2_{(10,\Sigma)}$						
	$\rho = 0.2$			$\rho = 0.7$		
n	**MC**	**FGLSR**	**FL**	**MC**	**FGLSR**	**FL**
15	0,046	0,039	0,026	0,041	0,041	0,027
35	0,438	0,416	0,379	0,405	0,296	0,060
65	0,935	0,940	0,911	0,927	0,900	0,172
100	1	1	0,998	1	0,999	0,341
150	1	1	1	1	1	0,570
200	1	1	1	1	1	0,754
500	1	1	1	1	1	1

Table VI.

Power properties for heteroscedastic disturbances, $\delta \sim \chi^2_{(10,\Sigma)}$				
	$\sigma_i^2 = X_{6,i}$		$\sigma_i^2 = 3X_{6,i} + 2X_{7,i}$	
n	**MC**	**FGLSR**	**MC**	**FGLSR**
15	0,047	0,031	0,050	0,032
35	0,433	0,424	0,430	0,421
65	0,935	0,940	0,932	0,944
100	0,999	0,999	1	0,999
150	1	1	1	1
200	1	1	1	1
500	1	1	1	1

Table VII.

Power properties for i.i.d. disturbances, $\delta \sim T_{(5,\Sigma)}$					
	Autocorrelation robust tests			Heteroscedasticity robust tests	
n	**MC**	**FGLSR**	**FL**	**MC**	**FGLSR**
15	0,032	0,026	0,019	0,031	0,021
35	0,370	0,354	0,376	0,364	0,342
65	0,753	0,741	0,758	0,749	0,741
100	0,927	0,919	0,933	0,924	0,927
150	0,990	0,999	0,992	0,990	0,990
200	1	1	1	1	1
500	1	1	1	1	1

Table VIII.

Power properties for autocorrelated disturbances, $\delta \sim T_{(5,\Sigma)}$						
	$\rho = 0.2$			$\rho = 0.7$		
n	**MC**	**FGLSR**	**FL**	**MC**	**FGLSR**	**FL**
15	0,033	0,027	0,023	0,033	0,041	0,026
35	0,368	0,360	0,344	0,358	0,287	0,094
65	0,750	0,744	0,719	0,735	0,712	0,263
100	0,952	0,921	0,907	0,923	0,912	0,405
150	0,990	0,990	0,984	0,990	0,990	0,576
200	1	1	1	1	1	0,705
500	1	1	1	1	1	0,973

Table IX.

Power properties for heteroscedastic disturbances, $\delta \sim T_{(5,\Sigma)}$				
	$\sigma_i^2 = X_{6,i}$		$\sigma_i^2 = 3X_{6,i} + 2X_{7,i}$	
n	**MC**	**FGLSR**	**MC**	**FGLSR**
15	0,036	0,021	0,033	0,023
35	0,373	0,360	0,379	0,346
65	0,751	0,734	0,749	0,740
100	0,926	0,925	0,925	0,924
150	0,991	0,991	0,991	0,989
200	1	1	1	1
500	1	1	1	1

VIII. Conclusions and summary

In this paper we have motivated why standard moment based tests for non-normality in general will not be valid for heteroscedastic or autocorrelated variables, in the sense that the type I error limits 100% if the data is not i.i.d. Hence, using regular non-normality tests on variables with a complicated data generating process, such as in economic applications, is questionable since a rejected null hypothesis may be due to non-normality *or* non-whiteness.

In order to overcome this problem we proposed a class of moment-based tests consistent to non-normality in presence of autocorrelation (i.e. *identically dependent distributed* disturbances) or heteroscedasticity (*independent heterogeneously distributed* disturbances) when applied to multivariate regression models.

The suggested tests rely on different levels of information of the data generating processes, namely (i) known covariance matrix (ii) known structure of the covariance matrix and (iii) no assumptions of the covariance matrix.

Since scalar covariance matrixes (i.e. $\mathbf{\Omega} = \mathbf{I}$) is a special case of a known covariance matrix, the class (i) test is more useful than it may seem, as scalar-valued covariance matrices do exist in some situations (indeed, this is the situation examined in most empirical studies of non-normality tests).

The test based on known structure of the covariance matrix is suitable in a FGLS framework while the non-normality test of class (iii) is useful in situations when normality is important and data are expected to be autocorrelated but nothing is known about the autocorrelation generating model.

A number of models were investigated in order to study the properties of the size, power and robustness of the tests. For each model we have performed 10 000 replications for various sample sizes ranging from 15 to 500 observations. In addition, the power properties have been examined for one skewed and one symmetric distribution.

The general findings are that the Monte Carlo tests hold the size exactly, as expected. The FGLSR tests slightly underestimate the size in small samples. The test of class (iii) also underestimates the size slightly. This effect carries over to the power properties in the sense that the class (i) tests have highest power, the class (ii) tests have lower power while the test of class (iii) has lowest power. Hence, the amount of information of the data generating process available has an immediate effect on the tests, the more information available, the better performance of the test. What is striking though is that, when the data is i.i.d., all our tests are well behaved, in the sense that the size is close to the nominal size and the power is high. Hence the loss in using our robust non-normality tests on i.i.d. data is minimal as compared to its classical non-robust counterparts. In other words, robustness is obtained without sacrificing efficiency.

Acknowledgement

The author is grateful to Ghazi Shukur and Max Petzold for valuable comments on earlier versions of the paper. Errors, of course, remain our sole responsibility.

Appendices A-G.

Appendix A.

Consider a normally distributed variate $\mathbf{x} \sim N\left(\mu \mathbf{1}_{(n\times 1)}, \mathbf{\Omega}_{(n\times n)}\right)$ where $\mu < \infty$ is a scalar and $\mathbf{\Omega}$ a p.d. matrix with typical elements $\omega_{ii} < \infty$, $\omega_{ii'} = 0$ for $i \neq i'$, and $(1/n)\sum_{i=1}^{n} \omega_i^j = \mathrm{O}(1)$ for $j = 1, 2, 3$. Further, let $\mathbf{1} := (1,...,1)'$ and $m^2 := (1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2 = (1/n)\mathbf{x}'\mathbf{A}\mathbf{x}$ where $\mathbf{A} := \mathbf{I}_n - (1/n)\mathbf{1}\mathbf{1}'$.

Proposition A.1.

(*a*) $E\left[m_2^2\right] = \left(1/n^2\right)\sum_{i\neq j}\sigma_i^2\sigma_j^2 + o\left(n^{-1/2}\right)$.

(*b*) $E\left[m_2^3\right] = \left(1/n^3\right)\sum_{i\neq j\neq k}\sigma_i^2\sigma_j^2\sigma_k^2 + o\left(n^{-1/2}\right)$.

Summations of the type $\sum_{i\neq j\neq k} a_i a_j a_k$ and $\sum_{i\neq j} a_i a_j$ contain $\left(n^3 - n\right)$ and $\left(n^2 - n\right)$ terms terms respectively (permutations of i, j and k thus being allowed).

Proof: Following Magnus (1978), the second and third moments of m_2 are

$$E\left[\left(1/n^2\right)\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)^2\right] = \left(1/n^2\right)\left\{\left(tr\left(\mathbf{A}\mathbf{\Omega}\right)\right)^2 + 2tr\left(\left(\mathbf{A}\mathbf{\Omega}\right)^2\right)\right\}.$$

$$E\left[\left(1/n^3\right)\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)^3\right] = \left(1/n^3\right)\left\{\left(tr\left(\mathbf{A}\mathbf{\Omega}\right)\right)^3 + 6tr\left(\mathbf{A}\mathbf{\Omega}\right)tr\left(\left(\mathbf{A}\mathbf{\Omega}\right)^2\right) + 8tr\left(\left(\mathbf{A}\mathbf{\Omega}\right)^3\right)\right\}.$$

Further, observe that each diagonal element of $\left(\mathbf{1}\mathbf{1}'\mathbf{\Omega}\right)^r$ is $\mathrm{O}(1)$ for $r = 1, 2, 3$, so

$tr\left\{\left((1/n)\mathbf{1}\mathbf{1}'\mathbf{\Omega}\right)^j\right\} = \left(1/n^j\right)n\mathrm{O}(1) = \left(1/n^{j-1}\right)\mathrm{O}(1)$ and hence

$$tr\left\{\left(\mathbf{A}\mathbf{\Omega}\right)^3\right\} = tr\left\{\left(\mathbf{\Omega}\right)^3 - 3\mathbf{\Omega}^2\left(1/n\right)\mathbf{1}\mathbf{1}'\mathbf{\Omega} + 3\mathbf{\Omega}\left(\left(1/n\right)\mathbf{1}\mathbf{1}'\mathbf{\Omega}\right)^2 - \left(\left(1/n\right)\mathbf{1}\mathbf{1}'\mathbf{\Omega}\right)^3\right\} =$$

$tr(\boldsymbol{\Omega})^3 - O(1) + O(n^{-1}) - O(n^{-2}) = tr\{(\boldsymbol{\Omega})^3 + O(1)\}$, and similarly,
$tr\{(\mathbf{A}\boldsymbol{\Omega})^2\} = tr\{(\boldsymbol{\Omega})^2 + O(1)\}$.

We then get

$$E[m_2^2] = (1/n^2)E[(\mathbf{x}'\mathbf{A}\mathbf{x})^2] = (1/n^2)\{(tr(\boldsymbol{\Omega}) + O(1))^2 + 2tr((\boldsymbol{\Omega})^2 + O(1))\} =$$
$$(1/n^2)\{(tr(\boldsymbol{\Omega}))^2 + 2O(1)tr(\boldsymbol{\Omega}) + (O(1))^2 + 2tr(\boldsymbol{\Omega})^2 + 2O(1)\} =$$
$$(1/n^2)\{(tr(\boldsymbol{\Omega}))^2 + O(n) + O(1) + O(n) + O(1)\} =$$
$$(1/n^2)(tr(\boldsymbol{\Omega}))^2 + o(n^{-1/2}) = (1/n^2)\left(\sum_{i=1}^n \sigma_i^2\right)^2 + o(n^{-1/2}) =$$
$$(1/n^2)\left(\sum_{i=1}^n \sigma_i^4\right) + (1/n^2)\left(\sum_{i\neq j} \sigma_i^2\sigma_j^2\right) + o(n^{-1/2}) = (1/n^2)\left(\sum_{i\neq j} \sigma_i^2\sigma_j^2\right) + o(n^{-1/2}).$$

Similarly,

$(1/n^3)E[(\mathbf{x}'\mathbf{A}\mathbf{x})^3] = (1/n^3)\{\Psi_1 + \Psi_2 + \Psi_3\}$, say, where

$$\Psi_1 = (tr(\boldsymbol{\Omega}) + O(1))^3 = (tr(\boldsymbol{\Omega}))^3 + 3(tr(\boldsymbol{\Omega}))^2 O(1) + 3(tr(\boldsymbol{\Omega}))(O(1))^2 + (O(1))^3 =$$
$$(tr(\boldsymbol{\Omega}))^3 + O(n^2) + O(n) + O(1) = (tr(\boldsymbol{\Omega}))^3 + O(n^2),$$

$$\Psi_2 = 6(tr(\boldsymbol{\Omega}) + O(1))(tr(\boldsymbol{\Omega}^2) + O(1)) = (O(n) + O(1))(O(n) + O(1)) = O(n^2),$$
and

$$\Psi_3 = 8(tr(\boldsymbol{\Omega})^3 + O(1)) = O(n) + O(1) = O(n).$$

Hence

$$(1/n^3)E[(\mathbf{x}'\mathbf{A}\mathbf{x})^3] = (1/n^3)\{(tr(\boldsymbol{\Omega}))^3 + O(n^2) + O(n)\} =$$
$$(1/n^3)\left(\sum_{i=1}^n \sigma_i^2\right)^3 + o(n^{-1/2}) = (1/n^3)\sum_{i=1}^n \sigma_{i=1}^6 + (1/n^3)\sum_{i\neq j\neq k} \sigma_i^2\sigma_j^2\sigma_k^2 + o(n^{-1/2})$$
$$= (1/n^3)\sum_{i\neq j\neq k} \sigma_i^2\sigma_j^2\sigma_k^2 + o(n^{-1/2}),$$ which completes Proposition A.1 ▣

Proposition A.2.

(*c*) $$E[m_3] = o\left(n^{-1/2}\right).$$

(*d*) $$E[m_4] = (3/n)\sum_{i=1}^{n}\sigma_i^4 + o\left(n^{-1/2}\right).$$

(*e*) $$E[m_3^2] = (1/n)\left\{(1/n)15\sum_{i=1}^{n}\sigma_i^6 + (9/n^3)\left(\sum_{i\neq j\neq k}\sigma_i^2\sigma_j^2\sigma_k^2\right) - (18/n^2)\sum_{i\neq j}\sigma_i^4\sigma_j^2\right\} + o\left(n^{-1}\right).$$

Proof: From Kendall and Stuart (1976) p. 245, we have

$$E[m_r] = E\left[\{1-r/n\}\sum x_i^r - (r/n)\sum_{i\neq j} x_i x_j^{r-1}\right]\Big/ n + o\left(n^{-1/2}\right), \text{ and}$$

$$E[m_r^2] = (1/n^2)E\left\{\sum_{i=1}^{n} x_i^{2r} + \sum_{i\neq j} x_i^r x_j^r + (r^2/n^2)\left(\sum_{i\neq j} x_i^r x_j^r + \sum_{i\neq j} x_i^2 x_j^{2r-2} + \sum_{i\neq j\neq k} x_i^2 x_j^{r-1} x_k^{r-1}\right) - (2r/n)\sum_{i\neq j} x_i^{r+1} x_j^{r-1}\right\} + \{E[m_r]\}^2.$$

Hence

$$E[m_3] = \{(1/n) - 3/n^2\}\sum E[x_i^3] - (3/n^2)\sum E[x_i]E[x_j^2] + o\left(n^{-1/2}\right) = o\left(n^{-1/2}\right).$$

$$E[m_4] = \{(1/n) - 4/n^2\}\sum E[x_i^4] - (4/n^2)\sum E[x_i]E[x_j^3] + o\left(n^{-1/2}\right) =$$
$$(1/n)\sum E[x_i^4] + o\left(n^{-1/2}\right) = (3/n)\sum\sigma_i^4 + o\left(n^{-1/2}\right).$$

$$E[m_3^2] = (1/n)\left\{(1/n)\sum E[x_i^6] + (1/n)\sum E[x_i^3]E[x_j^3] + (9/n^3)\sum E[x_i^3]E[x_j^3] + (9/n^3)\sum E[x_i^2]E[x_j^4] + (9/n^3)\sum E[x_i^2]E[x_j^2]E[x_k^2] - (6/n^2)\sum E[x_i^4]E[x_j^2]\right\} + \left\{o\left(n^{-1/2}\right)\right\}^2 =$$
$$(1/n)\left\{(1/n)15\sum_{i=1}^{n}\sigma_i^6 + (9/n^3)\left(\sum_{i\neq j\neq k}\sigma_i^2\sigma_j^2\sigma_k^2\right) - (18/n^2)\sum_{i\neq j}\sigma_i^4\sigma_j^2\right\} + o\left(n^{-1}\right),$$

and this completes proposition A.2 ▣

Let $x_i \sim N(\mu, \sigma_i^2)$ where $\sigma_i^2 = i/n$, and observe that $\sum_{i=1}^{n} i = (1/2)n(n+1)$ and $\sum_{i=1}^{n} i^2 = (1/6)n(n+1)(2n+1)$. Hence
$(1/n)\sum \sigma_i^2 = (1/n)\sum (i/n) = (1/2n^2)n(n+1) = O(1)$, and
$(1/n)\sum \sigma_i^4 = (1/n)\sum (i/n)^2 = (1/6n^3)\sum i^2 = (1/6n^3)n(n+1)(2n+1) = O(1)$.

In other words, the results of Propositions A1.*a* and A2.*d* hold for $\sigma_i^2 = i/n$, and we have the following:

Proposition A3.

$$\lim_{n\to\infty} \sqrt{n}\left(E(m_4)/E(m_2^2) - 3\right) = \infty .$$

Proof:

$$E[m_4] = (3/n)\sum \sigma_i^4 + o(n^{-1/2}) = (3/n)\sum (i/n)^2 + o(n^{-1/2}) = (3/n^3)\sum i^2 + o(n^{-1/2}) =$$
$$(3/n^3)(1/6)n(n+1)(2n+1) + o(n^{-1/2}) = (1/2)(1/n^2)(n+1)(2n+1) + o(n^{-1/2}).$$

$$E(m_2^2) = (1/n^2)\left(\sum \sigma_i^2\right)^2 + o(n^{-1/2}) = (1/n^2)\{(1/n)(1/2)n(n+1)\}^2 + o(n^{-1/2}) =$$
$$(1/4)(1/n^2)(n+1)^2 + o(n^{-1/2}).$$

Hence

$$\sqrt{n}\left(\{E(m_4)/E(m_2^2)\} - 3\right) \to \sqrt{n}\left(\{(1/2)(1/n^2)(n+1)(2n+1)/(1/4)(1/n^2)(n+1)^2\} - 3\right) =$$
$$= \sqrt{n}\left(\{2(2n+1)/(n+1)\} - 3\right) = O(n^{1/2}).$$

In other words, $\sqrt{n}\left(E(m_4)/E(m_2^2) - 3\right)$ is dominated by terms of order $O(n^{1/2})$ and so $\lim_{n\to\infty} \sqrt{n}\left(E(m_4)/E(m_2^2) - 3\right) = \infty$ ▣.

Appendix B. Limiting distribution of GLS residuals

The GLS residuals of the $g:th$ marginal model are $\hat{\tilde{\boldsymbol{\varepsilon}}}_{g,(n\times 1)} := \left(\tilde{\mathbf{Y}}_g - \hat{\tilde{\mathbf{Y}}}_{OLS,g}\right) = \left(\mathbf{I}_{(n\times n)} - \mathbf{H}\right)\tilde{\boldsymbol{\varepsilon}}_g$ for $\mathbf{H}_{(n\times n)} := \tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'$, where $\tilde{\mathbf{X}} = \boldsymbol{\Omega}^{-1/2}\mathbf{X}$, $\tilde{\boldsymbol{\varepsilon}}_g = \boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}_g$ and $\tilde{\mathbf{Y}}_g = \boldsymbol{\Omega}^{-1/2}\mathbf{Y}_g$, $g = 1,...,G$. Assuming that $V\left(\tilde{\boldsymbol{\varepsilon}}_g\right) = \sigma_g^2\mathbf{I}$, we get $V\left(\boldsymbol{\xi}_g\right) := V\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_g - \tilde{\boldsymbol{\varepsilon}}_g\right) = \sigma_g^2\mathbf{H}$, or elementwize, $V\left(\xi_{ig}\right) = V\left(\hat{\tilde{\varepsilon}}_{ig} - \tilde{\varepsilon}_{ig}\right) = \sigma_g^2 H_{ii}$, $i = 1,...,n$. By Chebychev's inequality then, $P\left(\left|\xi_{ig} - E\left(\xi_{ig}\right)\right| \geq \nu\right) = P\left(\left|\hat{\tilde{\varepsilon}}_{ig} - \tilde{\varepsilon}_{ig}\right| \geq \nu\right) \leq \frac{H_{ii}\sigma_g^2}{\nu^2}$. Thus $\max_{1\leq i\leq n}\left(H_{ii}\right) \to 0$ suffices for $\left|\hat{\tilde{\varepsilon}}_{ig} - \tilde{\varepsilon}_{ig}\right| \xrightarrow{p} 0$ to hold. Since the i:th diagonal element of $\mathbf{H}$ can be written as $H_{ii} = \left(\tilde{X}_{i1},...,\tilde{X}_{ik}\right)\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\left(\tilde{X}_{i1},...,\tilde{X}_{ik}\right)'$, and k is a fixed finite number, it follows that $plim\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} = \mathbf{0}$ suffices for $\max_{1\leq i\leq n}\left(H_{ii}\right) \to 0$ to hold. Now, $plim\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}/n\right)^{-1}$ is a finite non-singular matrix by our assumptions of (2.1), and as $\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} = (1/n)\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}/n\right)^{-1}$ it follows that $plim\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} = \mathbf{0}$. Finally we have from Rao (1973), p 122, $\left|\hat{\tilde{\varepsilon}}_{i,n} - \tilde{\varepsilon}_{i,n}\right| \xrightarrow{p} 0$, $\tilde{\varepsilon}_{i,n} \xrightarrow{p} \tilde{\varepsilon}_i \Rightarrow \hat{\tilde{\varepsilon}}_{i,n} \xrightarrow{\ell} \tilde{\varepsilon}_i$. Hence $\hat{\tilde{\varepsilon}}_{ig,n} \xrightarrow{\ell} \tilde{\varepsilon}_{ig}$.

Appendix C: Limiting distribution of FGLS residuals

According to Appendix B above, $\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS,i} - \tilde{\boldsymbol{\varepsilon}}_i \xrightarrow{p} \mathbf{0}$. Hence if $\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{\mathrm{FGLS},i} - \hat{\tilde{\boldsymbol{\varepsilon}}}_{\mathrm{GLS},i}\right) \xrightarrow{p} \mathbf{0}$, then $\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS,i}$ has the same asymptotic distribution as $\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS,i}$ (which is that of $\tilde{\boldsymbol{\varepsilon}}_i$). This property holds under a fairly wide class of situations. In particular, let $\boldsymbol{\Omega}_{(n\times n)} := \boldsymbol{\Omega}_{(n\times n)}\left(\boldsymbol{\theta}\right)$ and $\hat{\boldsymbol{\Omega}}_{(n\times n)} = \boldsymbol{\Omega}_{(n\times n)}\left(\hat{\boldsymbol{\theta}}\right)$ where $\boldsymbol{\theta}$ is a parameter of finite dimension (and this dimension does not depend on n) such that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$. In other words, $\boldsymbol{\Omega}$ is assumed to be a function of a finite-dimension parameter to which there exists a corresponding consistent estimate. Then, under most familiar settings met, we have $\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{\mathrm{FGLS},i} - \hat{\tilde{\boldsymbol{\varepsilon}}}_{\mathrm{GLS},i}\right) \xrightarrow{p} \mathbf{0}$. Of cause, this is something to be checked in every case. Below

we provide an example of how such a proof may be given: Let $\varepsilon_{ig} = \phi\varepsilon_{i-1,g} + \delta_{ig}$ where $\delta_{ig} \sim iid\left(0,\sigma^2\right)$. Then $\left(E\left[\boldsymbol{\varepsilon}_g\boldsymbol{\varepsilon}_g'\right]\right)^{-1} = \left(1/\sigma^2\right)\boldsymbol{\Omega}^{-1} = \left(1/\sigma^2\right)\mathbf{P}'\mathbf{P}$ where

$$\mathbf{P} = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 \\ -\phi & 1 & 0 & \cdots & 0 \\ 0 & -\phi & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \\ 0 & & & -\phi & 0 \end{bmatrix} \tag{C.1}$$

with its corresponding estimate $\hat{\mathbf{P}} = \mathbf{P}\left(\hat{\phi}\right)$, i.e. (C.1) with ϕ replaced by $\hat{\phi}$.

Hence $\hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS,g} = \tilde{\mathbf{Y}}_g - \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}_g = \boldsymbol{\Omega}^{-1/2}\left(\mathbf{Y}_g - \mathbf{X}\hat{\boldsymbol{\beta}}_g\right) = \mathbf{P}\left(\mathbf{Y}_g - \mathbf{X}\hat{\boldsymbol{\beta}}_g\right)$ and so $\left(\hat{\tilde{\boldsymbol{\varepsilon}}}_{FGLS,g} - \hat{\tilde{\boldsymbol{\varepsilon}}}_{GLS,g}\right) = \left(\hat{\mathbf{P}} - \mathbf{P}\right)\left(\mathbf{Y}_g - \mathbf{X}\hat{\boldsymbol{\beta}}_g\right)$. The first element of this vector becomes $\left(\hat{\tilde{\varepsilon}}_{FGLS,1g} - \hat{\tilde{\varepsilon}}_{GLS,1g}\right) = \left\{\left(\sqrt{1-\hat{\phi}^2} \quad 0 \quad \cdots \quad 0\right) - \left(\sqrt{1-\phi^2} \quad 0 \quad \cdots \quad 0\right)\right\}\left(Y_{1g} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_g\right)$, the second $\left(\hat{\tilde{\varepsilon}}_{FGLS,2g} - \hat{\tilde{\varepsilon}}_{GLS,2g}\right) = \left\{\left(-\hat{\phi} \quad 1 \quad \cdots \quad 0\right) - \left(-\phi \quad 1 \quad \cdots \quad 0\right)\right\}\left(Y_{2g} - \mathbf{X}_2\hat{\boldsymbol{\beta}}_g\right)$, and so on. Hence, if $\hat{\phi} - \phi \xrightarrow{p} 0$, then $\left(\hat{\tilde{\varepsilon}}_{FGLS,ig} - \hat{\tilde{\varepsilon}}_{GLS,ig}\right) \xrightarrow{p} 0$. Similar proofs may be given for general ARMA(p,q) processes, the elements of the corresponding $\mathbf{P}$ matrix are given in Gailbraith and Zinde-Walsh (1992). Furthermore, note that the inverse matrix root of a diagonal matrix, such as that of a heteroscedastic model where $V\left[\varepsilon_{ig}\right] = \omega_{ig} = \theta_1 + \theta_2 x_i$, say, is $\omega_{ig}^{-1/2}$ on its diagonal and zero elsewhere. Hence, if there exists a consistent estimate of ω_{ig} (e.g. $\hat{\omega}_{ig} = \hat{\theta}_1 + \hat{\theta}_2 x_i$ for an observable x_i and $\hat{\theta}_j \xrightarrow{p} \theta_j$, $j = 1,2$) then our example above covers this situation as well. However, again, this is something to be checked in every case.

Appendix D.

Consider the (transposed) OLS residual matrix $\hat{\boldsymbol{\varepsilon}}_{(G\times n)} = \boldsymbol{\varepsilon}_{(G\times n)}\mathbf{M}_{(n\times n)}$, $\mathbf{M} = \mathbf{I}_{(n\times n)} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, and the corresponding sample covariance matrix $\mathbf{S}_{\hat{\boldsymbol{\varepsilon}}} := \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}'/n = \mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}'$, say, where $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}$ is the Cholesky root of $\mathbf{S}_{\hat{\boldsymbol{\varepsilon}}}$. Next, consider a scale-transformed disturbance variable $\boldsymbol{\delta}_{(G\times n)} := \boldsymbol{\Sigma}^{1/2}_{(G\times G)}\boldsymbol{\varepsilon}$ and its observable counterpart $\hat{\boldsymbol{\delta}}_{(G\times n)} = \boldsymbol{\delta}\mathbf{M} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\varepsilon}\mathbf{M}$. The corresponding covariance matrix may then be written as $\mathbf{S}_{\hat{\boldsymbol{\delta}}} := \hat{\boldsymbol{\delta}}\hat{\boldsymbol{\delta}}'/n = \boldsymbol{\Sigma}^{1/2}\left(\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\varepsilon}}'/n\right)\boldsymbol{\Sigma}'^{1/2} = \boldsymbol{\Sigma}^{1/2}\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}'\boldsymbol{\Sigma}'^{1/2} = \boldsymbol{\Sigma}^{1/2}\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}\right)' = \mathbf{L}_{\hat{\boldsymbol{\delta}}}\mathbf{L}_{\hat{\boldsymbol{\delta}}}'$, say.

Then we have

$$\mathbf{L}_{\hat{\boldsymbol{\delta}}}^{-1}\hat{\boldsymbol{\delta}} = \left(\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\boldsymbol{\Sigma}^{-1/2}\right)\left(\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{\varepsilon}}\right) = \mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}, \tag{D1}$$

i.e. the "studentized" residual matrix $\mathbf{L}_{\hat{\boldsymbol{\delta}}}^{-1}\hat{\boldsymbol{\delta}}$ is identical to that of $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}$. In other words, $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}$ does not depend on $\boldsymbol{\Sigma}$. In particular, if $\boldsymbol{\delta}_i \sim N(\mathbf{0},\mathbf{I})$ and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0},\boldsymbol{\Sigma})$, then $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}} \sim \mathbf{L}_{\hat{\boldsymbol{\delta}}}^{-1}\hat{\boldsymbol{\delta}}$. However, the distribution of the scaled residuals $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}$ depend on the $\mathbf{M}$ matrix, since $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}\mathbf{M}$. But as $\mathbf{M}$ is a fixed, observable matrix, we can generate an arbitrary number of independent Monte Carlo realisations of $\boldsymbol{\varepsilon}$, $\{\boldsymbol{\varepsilon}_b\}_{b=1}^{B}$, say, and from each drawing form $\boldsymbol{\varepsilon}_b \mapsto \boldsymbol{\varepsilon}_b\mathbf{M} = \hat{\boldsymbol{\varepsilon}}_b$ and then $\hat{\boldsymbol{\varepsilon}}_b \mapsto \mathbf{L}_{\hat{\boldsymbol{\varepsilon}}_b}^{-1}\hat{\boldsymbol{\varepsilon}}_b$. Hence, any bounded function of $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}_b}^{-1}\hat{\boldsymbol{\varepsilon}}_b$ will have the same distribution as $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}$. In particular, if the statistic (5.1) on p. 10 is calculated from the Monte Carlo drawing $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}_b}^{-1}\hat{\boldsymbol{\varepsilon}}_b$, it will have the same distribution as if calculated from $\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}$, i.e. $\tilde{b}\left(\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}}^{-1}\hat{\boldsymbol{\varepsilon}}\right) \sim \tilde{b}\left(\mathbf{L}_{\hat{\boldsymbol{\varepsilon}}_b}^{-1}\hat{\boldsymbol{\varepsilon}}_b\right)$. ▣

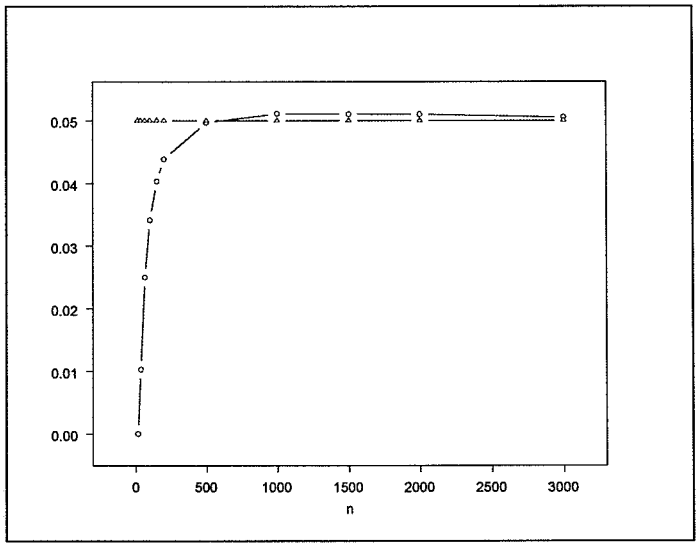
Appendix E.

Let $\{x_i\}$ be the stationary process $x_i = \sum_{k=0}^{\infty} \psi_k z_{i-k}$, $\{z_i\} \sim iid(0, \sigma^2)$, where $0 < \sigma^2 < \infty$, $\sum_{k=0}^{\infty} |\psi_k| < \infty$ and $E[z_i^4] < \infty$. Then let $\gamma_k := E[x_i x_{i+k}]$, $\gamma_k = \gamma_{-k}$ and $\rho_k := \gamma_k / \gamma_0$, where $0 < \gamma_0 < \infty$ and $|\gamma_k| < \gamma_0$ $(k > 1)$ by assumption. Now, consider the sums $S_3 := \sum_{k=-\infty}^{k=\infty} \rho_k^3$ and $S_4 := \sum_{k=-\infty}^{k=\infty} \rho_k^4$. We here establish the existence of consistent estimates of S_3 and S_4:

Firstly, note that $\sum_{k=0}^{\infty} |\psi_k| < \infty \Rightarrow \sum_{k=0}^{\infty} |\gamma_k| < \infty$ (Hamilton (1994) p. 52). From our assumptions above we then have $(1/\gamma_0) \sum_{k=0}^{\infty} |\gamma_k| = \sum_{k=0}^{\infty} |\gamma_k / \gamma_0| = \sum_{k=0}^{\infty} |\rho_k| < \infty$. Further, as $|\rho_k| < 1$ for $k > 1$ it follows that $|\rho_k^d| < |\rho_k|$ $d \in \{3,4\}$ and hence $\sum_{k=0}^{\infty} |\rho_k^d| < \infty$ and so there exists a $q < \infty$ such that $|\rho_{q+1}^d| + |\rho_{q+2}^d| + ... < \delta$ holds for any $\delta > 0$. In other words, $\sum_{k=1}^{\infty} \rho_k^d$ may be approximated by $\sum_{k=1}^{q} \rho_k^d$ for sufficiently large q. Now, consider the sample autocorrelation $\hat{\rho}_{k,n} = \hat{\gamma}_{k,n} / \hat{\gamma}_{0,n}$ where $\hat{\gamma}_{k,n} = (1/n) \sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})$. Then for each $k \in \{1, 2, ...\}$ we have $\hat{\rho}_{k,n} \xrightarrow{\ell} N(\rho_k, n^{-1} W_k)$ for some $W_k < \infty$ (Brockwell and Davis (1991), p. 221). Then $V[n^{4/10} \hat{\rho}_{k,n}] = n^{8/10} V[\hat{\rho}_{k,n}] = n^{-1/5} W_k \to 0$ and so $(\hat{\rho}_{k,n} - \rho_k) = o(n^{-4/10})$. Furthermore, $(\hat{\rho}_{k,n}^3 - \rho_k^3) = (\hat{\rho}_{k,n} - \rho_k)(\hat{\rho}_{k,n}^2 + \hat{\rho}_{k,n} \rho_k + \rho_k^2) = o(n^{-4/10}) O(1) = o(n^{-4/10})$, and $(\hat{\rho}_k^4 - \rho_k^4) = (\hat{\rho}_k - \rho_k)(\hat{\rho}_k + \rho_k)(\hat{\rho}_k^2 + \rho_k^2) = o(n^{-4/10}) O(1) = o(n^{-4/10})$. Finally, consider the sum $\Delta_q^3 := \sum_{k=1}^{q} (\hat{\rho}_{k,n}^3 - \rho_k^3)$. Then Δ_q^3 contains q terms of order $o(n^{-4/10})$, so if we chose $q \le n^{4/10}$, it follows that $\sum_{k=1}^{q} \hat{\rho}_{k,n}^3 \xrightarrow{p} \sum_{k=1}^{q} \rho_k^3$, and as $\rho_{-k} = \rho_k$ and $\rho_0^3 \equiv 1$ we get $1 + 2\sum_{k=1}^{q} \hat{\rho}_{k,n}^3 \xrightarrow{p} \sum_{k=-q}^{q} \rho_k^3$ and similarly, $1 + 2\sum_{k=1}^{q} \hat{\rho}_{k,n}^4 \xrightarrow{p} \sum_{k=-q}^{q} \rho_k^4$ ▣.

Appendix F.

Simulated true size for the Lütkepohl and Theilen (1991) test for $\alpha = 5\%$ nominal size using 1 million replications.



Simulated critical values for the Lütkepohl and Theilen (1991) test using 1 million replications.

Empirical critical values	Sample size
13.81464	15
14.49681	35
15.80272	65
16.76972	100
17.41341	150
17.76464	200
18.28421	500
18.38602	1000
18.38040	1500
18.37508	2000
18.34429	3000
18.30704	∞

References

Bewley, R. (1986). Allocation models, Specification, Estimation and Applications, Ballinger Publishing Company.

Bickel, P. J. (1978). "Tests for Heteroscedasticity, Nonlinearity and Nonadditivity." *Annals of Statistics* 6: 266-291.

Breusch, T. S. (1978). "Testing for Autocorrellation in Dynamic Linear Models." *Australian Economic Papers* 17: 334-355.

Brockwell, P. J. and R. A. Davis (1991). Time Series: Theory and Methods, Springer.

Dufour, J. M., F. Abdeljelil, et al. (1998). "Simulation-based finite sample normality tests in linear regressions." *Econometrics Journal* 1: 154-173.

Gailbraith, J. and V. Zinde-Walsh (1992). "The GLS Transformation Matrix and a Semi-Recursive Estimator for the Linear Regression Model With ARMA Errors." *Econometric Theory* 8.

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.

Holgersson, T. and G. Shukur (2001). "Some Aspects of Non-Normality Tests in Systems of Regression Equations." *Commun stat-simul* 30(2): 291-310.

Horswell, R. L. and S. Looney, W. (1993). "Diagnostic Limitations of Skewness Coefficients in assessing Departures from Univariate and Multivariate Normality." *Commun. Statist._Simula.* 22(2): 437-459.

Jarque, C. M. and A. K. Bera (1987). "A Test for Normality of Observations and Regression Residuals." *International Statistical Review* 55: 163-172.

Jarque, C. M. and C. R. Mckenzie (1995). "Testing for Multivariate Normality in Simultanous-Equations Models." *Mathematics and Computers in Simulation* 39(3-4): 323-328.

Kendall, M. and A. Stuart (1976). The Advanced Theory of Statistic, Volume 1, Distribution Theory. London, Charles Griffin&Company Limited.

Koenker, R. and G. Basset (1982). "Robust tests for Heteroscedasticity Based on Regression Quantiles." *Econometrica* 50(1): 43-62.

Lomnicki, Z. A. (1961). "Test for Departure from Normality in the Case of Linear Stochastic Processes." *Metrica* 4: 37-62.

Lütkepohl, H. and W. Schneider (1989). "Testing for Nonormality of Autoregressive Time Series." *Computational Statistics Quarterly* 2: 151-168.

Lütkepohl, H. and B. Theilen (1991). "Meassures of multivariate skewness and kurtosis for tests of nonormality." *Statistical Papers* 32: 179-193.

Magnus, J. (1978). "The Moments of products of quadratic Forms in Normal variables." *Statistica Neerlandica* 34(4): 201-210.

Malkovich, J. F. and A. A. Afifi (1973). "On Tests For Multivariate Normality." *J. Am. Statist. Assoc.* 68: 176-179.

Rao, C. R. (1973). Linear Statistical Inference. New York, Wiley.

Shukur, G. (2000). "The robustness of the systemwise Breusch-Godfrey autocorrelation test for non-normal distributed error terms." *Commun stat-simul* 29(2): 419-448.

Srivastava, M. S. a. K., C.G. (1979). An Introduction to Multivariate Statistics. New York, North Holland.

Stuart, A. and J. K. Ord (1993). "Kendalls's Advanced Theory of Statistics, Vol 1, Distribution theory."

Research Report

2001:5	Andersson, E., Bock, D. & Frisén, M.:	Likelihood based methods for detection of turning points in business cycles. A comparative study.
2001:6	Sonesson, C.:	Evaluations of some exponentially weighted moving average methods.
2001:7	Sonesson, C.:	Statistical surveillance. Exponentially weighted moving average methods and public health monitoring.
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. & Frisén, M.:	Detection of turning points in business cycles.
2002:7	Andersson, E., Bock, D. & Frisén, M.:	Some statistical aspects on methods for detection of turning points in business cycles.
2002:8	Andersson, E., Bock, D. & Frisén, M.:	Statistical surveillance of cyclical processes with application to turns in business cycles.