Research Report
Department of Statistics
Göteborg University
Sweden

On assessing multivariate normality

H.E.T. Holgersson

On assessing multivariate normality

By
H.E.T. Holgersson

Department of Statistics,
School of Economics and Commercial Law, Göteborg University,
Box 660 SE-405 30 Göteborg, Sweden.

Statistical analysis frequently relies on the assumption of normality. Though normality may often be relaxed in view of inferences of for example population expectations, it can be crucial in other aspects such as diagnostic tests or prediction intervals. It is then important to apply a hypothesis test against possible non-normality. But as the normality assumption usually regards normality of an unobservable variable, the test has to be applied on an observable proxy variable instead (usually the residuals), which may invoke biases in small samples. Additional problems arise as most tests for non-normality are valid only if the variables are independently and identically distributed (iid), a property often violated in for example economic applications.

This thesis consists of two papers dealing with the properties of non-normality tests in multivariate regression models. We give here a brief summary of the contents of the two papers.

The first paper, (written jointly with Ghazi Shukur), gives a short background of an omnibus test against non-normal multivariate skewness and kurtosis, namely the Jarque&McKenzie test. The small sample properties of the test are examined in view of robustness, size and power when applied to OLS residuals from systems of regression equations. The investigation has been performed using Monte Carlo simulations where factors like e.g. the number of equations, nominal sizes and degrees of freedom have been varied. Our analysis reveals four factors that have a bearing on the performance of the JM test's nominal size when applied to residuals, namely the degrees of freedom, number of equations, autocorrelation and distribution of regressors. Especially, we show that autocorrelation will ruin the test completely, in the sense that the true size will limit one, no matter the nominal size. Moreover, we show that a simple transformation of the residuals along with empirical critical values

will provide exact size regardless of distribution of regressors, number of degrees of freedom or number of equations, as long as the variables are iid. The power of the test is examined using heavy-tailed distributions. In general, the test has high power against the alternative distributions examined. In stark contrast, the power has shown to be zero for independent marginal distributions with normal skewness and kurtosis.

The second paper concerns the problem of testing for non-normality in multivariate models with nonspherical disturbances. We give an explicit reason why moment based non-normality tests, such as the popular Jarque&Bera test and multivariate extensions, in general fails if the variables are not iid. We propose several possible choices of proxy variables to the unobservable errors, which are applicable to non-normality testing as long as the structure of the covariance matrix is known. However, we show by Monte Carlo simulations that even a small misspecification of the covariance structure may well lead to an inconsistent test procedure, in the sense that the size will limit unity. Thus, the use of regular non-normality tests on variables with a complicated data generating process, such as in economic applications, is dubious. In addition our simulations reveal that the power can be reduced if the covariance matrix is unknown.

In all, the two papers concern the problem of assessing normality on unobservable multivariate variables. The properties of the test methods have been investigated with respect to size and power under conditions that are of relevance in empirical studies. We have also proposed methods for controlling the size when the covariance structure is known. Moreover, as opposed to many other inference procedures where a good approximation of the covariance suffices to provide sound results, we conclude that non-normality testing must be done with great care.

Some Aspects of Non-Normality Tests
in Systems of Regression Equations

H.E.T. Holgersson and G. Shukur

Department of Statistics
Göteborg University
SE- 405 30 Göteborg
Sweden

ABSTRACT

In this paper, a short background of the Jarque and McKenzie (JM) test for non-normality is given, and the small sample properties of the test is examined in view of robustness, size and power. The investigation has been performed using Monte Carlo simulations where factors like, e.g., the number of equations, nominal sizes, degrees of freedom, have been varied.

Generally, the JM test has shown to have good power properties. The estimated size due to the asymptotic distribution is not very encouraging though. The slow rate of convergence to its asymptotic distribution suggests that empirical critical values should be used in small samples.

In addition, the experiment shows that the properties of the JM test may be disastrous when the disturbances are autocorrelated. Moreover, the simulations show that the distribution of the regressors may also have a substantial impact on the test, and that homogenised OLS residuals should be used when testing for non-normality in small samples.

Key words : Non-normality test; Systems of equations; Residuals; Monte Carlo.

JEL Classification: C32

I. INTRODUCTION

The main purpose of this paper is to investigate the small sample performances of the Jarque and McKenzie (JM) test for non-normality when applied to system of regression equations, in view of robustness, size and power.

The normal distribution is often considered as a mathematical abstraction without connection to reality. Some scientists even state that normality is a pure myth (e.g.,Geary (1)). However, these claims are often based on the bare fact that an observable random variable X_d (d being the dimension of the variable) rarely fulfils two fundamental properties of the normal distribution; namely that the sample space should equal $\mathbb{R}^d$, (for example, a one-dimensional random variable should be defined on the whole real line), and that of symmetry. Thus, variables such as the weight of newborn babies or the number of sunspots per day can never be normally distributed. Yet, these variables can often be *approximately* normally distributed, in the sense that the normal-theory can be used on them resulting in reasonable inferences of their nature.

In this paper, however, we will approach the theory of normality from a different point of view; many stochastically phenomenons are assumed generated by one deterministic component and one stochastic, the latter being an unobservable error term. This random component is much more in line with the normal theory than is the observable ones. For example, the two conditions above mentioned are intuitively satisfied when the random variable is defined as deviations from a certain central measure.

Particularly, in regression analysis the error terms are frequently assumed to be normally distributed. It is not very likely that small deviations from this assumption will cause any serious inferential complications. On the other hand, when the deviations are large it is well known that diagnostic test based on the estimated versions of the disturbances will be suspect. Therefore, it is crucial that the distributional properties of the disturbances are examined carefully.

The history of normality test goes way back to the early century. The first well-known test is probably that of Kolmogorov (2), who suggested a (non-parametric) test using the empirical distribution function. The topic has then been developed successively, and a large variety of methods have been proposed, both for univariate and multivariate variables. Yet, it seems like minimal research has been made on the empirical properties on the latter of these methods.

Apart from this, we have the additional complication of handling an unobservable random variable, rather than an observable. Habitually, the so-called OLS-residuals are used as a proxy to the (possible multivariate) unobservable variable. This method may seem natural and intuitive as the residuals (under regular premises) are consistent estimates of the true errors.

It has been shown that the distribution of any goodness-of-fit statistic, which depends only on the empirical distributions of the residuals, converges to that of the true variable (e.g., Pierce and Kopecky (3)). Unfortunately, asymptotic results like this has implicated that residuals often are used as if they were identical to the disturbances. Consequently, the small sample properties of this negligent use have been given brief attention.

The paper is organised as follows: Section II discuss the model specification. In Section III, we discuss non-normality test, while in Section IV we present our Monte Carlo design used in this paper. In section V we present our most interesting results regarding the simulations. Finally, in section VI we give a brief summarisation.

II. MODEL SPECIFICATION

In this section, we will set up some standard assumptions of the underlying model. We do not claim that these are always realistic, but they do provide an idea of how the non-normality test that we are about to examine behaves under idealistic situations.

Consider a standard linear regression model

$$\mathbf{Y}_{(n\times 1)} = \mathbf{X}_{(n\times k)}\boldsymbol{\beta}_{(k\times 1)} + \boldsymbol{\varepsilon}_{(n\times 1)} \tag{1}$$

so that the model contains k parameters, with $\mathbf{X}$ strictly exogenous. Especially, the first column of $\mathbf{X}$ contains a unit vector. We will then make the following assumptions:

i. $|\mathbf{X}'\mathbf{X}| \neq 0$ ($\mathbf{X}$ is of full rank)

ii. $V(\boldsymbol{\varepsilon}|\mathbf{X}) = \sigma^2 \mathbf{I}$

iii. $E(\boldsymbol{\varepsilon}|\mathbf{X}) = \mathbf{0}$

iv. $\lim\limits_{n\to\infty}\left(\dfrac{1}{n}\mathbf{X}'\mathbf{X}\right)^{-1} = \mathbf{Q}^{-1}$, a finite matrix .

v. $f(\boldsymbol{\varepsilon}) = f(-\boldsymbol{\varepsilon})$ (where f is the density function).

Assumption i. is not crucial; it serves merely to simplify the calculations. On the contrary, assumption iv. is of great importance in the asymptotic theory of regression analysis. Whenever this limit is not a finite matrix, the point estimates of the regression parameters may not be consistent, in which case the residuals will not be consistent estimates of the true disturbances. It should be noted that the regressors are usually stochastic in economic data, meaning that all statistics in this paper based on residuals, will contain a stochastic component, C say, that will be an ancillary statistic. However, we can treat the regressors as if they were fix by conditioning on C (Cox and Hinkley (4)) as long as assumption (iii) holds. This will implicitly be made throughout the paper. The last assumption is basically enforced in order to restrict the study within reasonable bounds.

The above analysis is, however, only strictly applicable in a single equation environment. Many models are expressed in terms of systems of equations, for example time series models across different units, in particular demand and production functions. In general, some sort of covariance structure will connect the models.

Treating each model separately, and performing a succession of single equation misspecification tests, will lead to the problem of mass significance. Even though this problem can be handled by using multiple inference, e.g. the Bonferoni inequality or the union-intersection method, this would lead to a reduction in the validity of our conclusions as the problem is in its very nature multivariate.[*] Therefore it is necessary to consider several models jointly in a multivariate model. In this paper, we will limit ourselves to the simplest models.

Consider a system of linear regression equations

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

The residuals of the system are defined as

$$
\hat{\boldsymbol{\varepsilon}}_{(np\times1)} = \begin{bmatrix} \hat{\boldsymbol{\varepsilon}}_1 \\ \hat{\boldsymbol{\varepsilon}}_2 \\ \vdots \\ \hat{\boldsymbol{\varepsilon}}_P \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 \\ \mathbf{Y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 \\ \vdots \\ \mathbf{Y}_P - \mathbf{X}_P\hat{\boldsymbol{\beta}}_P \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1\boldsymbol{\varepsilon}_1 \\ \mathbf{M}_2\boldsymbol{\varepsilon}_2 \\ \vdots \\ \mathbf{M}_P\boldsymbol{\varepsilon}_P \end{bmatrix}, \text{ where } \mathbf{M}_i = \mathbf{I} - \mathbf{X}_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'.
$$

Throughout this paper we will assume that $\mathbf{M}_1 = \mathbf{M}_2 = \dots = \mathbf{M}_p = \mathbf{M}$.

Since $\mathbf{M}$ is symmetrical, there exists an $\mathbf{L}$ such that $\mathbf{M} = \mathbf{L}_{(n\times q)}\mathbf{L}_{(q\times n)}'$, where q is the rank of $\mathbf{M}$ which equals $(n\text{-}k)$. As $\mathbf{M}$ is idempotent we have $\mathbf{L}\mathbf{L}'\mathbf{L}\mathbf{L}' = \mathbf{L}\mathbf{L}' \Rightarrow \mathbf{L}'\mathbf{L} = \mathbf{I}_{(q\times q)}$. Then, if we define $\tilde{\boldsymbol{\varepsilon}} = \mathbf{L}'\boldsymbol{\varepsilon}$ it follows that $\tilde{\boldsymbol{\varepsilon}} \sim \mathbf{N}\left(0, \sigma^2\mathbf{I}_{q\times q}\right)$. These residuals can be considered as homogenised OLS residuals, and in what follows will be referred to as HOLS.

---

[*] *Edgerton et. al. (1996) and Shukur (1997) argue strongly for the use of systemwise misspecification tests.*

There may be other useful residuals than these mentioned above (e.g., Theil's BLUS residuals (Theil (5)) or stepwise residuals (Hedayat and Robson (6), Brown et.al. (7)). However, because of the simple structure and the ease of interpretation of OLS residuals and HOLS residuals, the study will be limited to these two types.

III. NON-NORMALITY TESTS

When testing for a particular distribution (or rather, for the deviation from an assumed distribution), it may seem natural to consider what characterise this certain distribution. The normal distribution is characterised by many features, in the sense that it possess properties that are unique for its distribution (e.g., Bryc (8), Lucaks and Laha (9)). Consequently, many of these properties have been used to test for normality. For example, the normal distribution maximises entropy against any other distribution with the same variance (e.g., Vasicek (10)), $\overline{X}$ and S_X^2 are independent iff X follow a normal distribution (e.g., Rao (11)), and so on. Then, there exist characterisations unique for *any* distribution, such as for example the distribution function and the characteristic function. Empirical versions of these have been used to test for non-normality as well (e.g., Kolmogorov (2), Epps (12)). Conversely, it is not generally true that the *moments* of a distribution uniquely determine the distribution of a random variable. It is well known that it is possible to find two distinct distribution functions that have the same set of moments (e.g., Heyde (13)). A sufficient condition for the moment sequence $\{\mu_k\}$ of a random variable X to uniquely determine the density function of X, is that the series $\sum_{k=1}^{n} \frac{\mu_k}{k!} s^k$ converges absolutely for some s > 0. That this is indeed true for the normal distribution is well known (e.g., Bryc (8)). Consequently, it is possible to construct useful statistics that are based on functions of moments to test for non-normality. For example, moment ratios are defined as

$$\beta_{2n+1} = \frac{\mu_3 \mu_{2n+3}}{\mu_2^{n+3}}, \qquad \beta_{2n} = \frac{\mu_{2n+2}}{\mu_2^{n+1}}. \qquad (2)$$

Especially, we have the well-known quantities

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}}, \qquad \gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3. \qquad (3)$$

which are the skewness and kurtosis respectively (e.g., Kendall and Stuart (14)). It is important to mention that tests based on skewness coefficients do not reliably discriminate between skewed and non-skewed distributions. This has been noted by several authors, e.g. Rayner and Best (15), who concludes that "moment ratios are not useful for the diagnosis of the type of non-normality". Horswell and Looney (16) writes, "The use of skewness tests to discriminate between skewed and symmetric distributions lacks theoretical foundation". Churchhill (17) proved, by giving a counterexample, that a distribution need not be symmetric even though all its odd moments vanish.

Even though these arguments do not necessarily imply that moment ratios are strictly non-diagnostic in all possible situations, we choose to focus on an over-all (or "omnibus") test instead. Jarque and Bera (18) suggested such an omnibus test for non-normality by considering a density function of the Pearson family

$$\frac{\partial f(x_i)}{\partial x_i} = (c_1 - x_i) f(x_i) / (c_0 - c_1 x_i + c_2 x_i^2),$$

and specifying the hypothesis $H_0 : c_1 = c_2 = 0$ (x_i is normally distributed).

By using the Lagrange multiplier approach they suggested the well-known statistic

$$T = N \left(\frac{\hat{\gamma}_1^2}{6} + \frac{(\hat{\gamma}_2 - 3)^2}{24} \right), \quad \text{where N is the number of observations.} \qquad (4)$$

The statistic T (some times denoted as JB) is asymptotically distributed $\chi^2_{(2)}$ under the null hypothesis.

Statistics expressing multivariate skewness and kurtosis have been proposed by several authors, e.g. Mardia (19), Malkovich and Afifi (20) and Srivastava (21). The first of these is defined as follows:

$$\hat{\gamma}_{1,p}(\mathbf{x}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}^3}{n^2}, \quad \text{and} \quad \hat{\gamma}_{2,p}(\mathbf{x}) = \frac{\sum_{i=1}^{n} r_i^4}{n}, \tag{5}$$

where

$$r_i^2 = \left[(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})\right]^2, \quad r_{ij} = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \quad \text{and} \quad \mathbf{S} = \mathbf{x}'\mathbf{x} - \bar{\mathbf{x}}\bar{\mathbf{x}}'.$$

The population counterparts of (5) are

$$\gamma_{1,p} = E\left[(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)\right]^3, \quad \gamma_{2,p} = E\left[(\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu)\right]^2. \tag{6}$$

For a location scale variable $\mathbf{X}_p$, we have the well known results

$$D_1(\mathbf{x}) = \frac{N\hat{\gamma}_{1p}}{6} \overset{L}{\sim} \chi^2_{(p(p+1)(p+2)/6)}, \quad D_2(\mathbf{x}) = \frac{\hat{\gamma}_{2p} - P(P+2)}{\sqrt{8NP(P+2)}} \overset{L}{\sim} N(0,1).$$

Jarque and Mckenzie, (22) suggested the combination

$$D_p(\mathbf{x}) = D_1(\mathbf{x}) + D_2^2(\mathbf{x}) \quad \text{with} \quad D_p(\mathbf{x}) \overset{L}{\sim} \chi^2_{1+p(p+1)(p+2)/6}. \tag{7}$$ *

The null hypothesis $H_0 : \boldsymbol{\varepsilon} \sim N(0, \Sigma)$ is then rejected at the α-level whenever $D_p(\boldsymbol{\varepsilon}) > \eta$ where $P(D_p(\boldsymbol{\varepsilon}) > \eta \mid \gamma_{1,P} = 0 \cap \gamma_{2,P} = P(2+P)) = \alpha$. This test will be the focus of the paper. It can be shown that $D_p(\hat{\boldsymbol{\varepsilon}})$ have the same asymptotic distribution as $D_p(\boldsymbol{\varepsilon})$. It should be noted though, that the estimated skewness and kurtosis are not unbiased. Huang and Bolch (23) showed that the skewness and kurtosis of the residuals are always biased towards their expected values under the H_0 when H_0 is false.

---

* Note that when P=1 the D_p statistic reduces to the JB statistic in (4).

A nice fact is that D_p is invariant to linear transformations. This compensates the well-known fact that D_1 and D_2^2 converge very slowly to its asymptotic distributions, since modern computers provide us with the possibility of simulating empirical critical values with high precision. All together, we find the JM test suitable for systems of linear regression models.

There are several other omnibus tests for normality that have been shown to perform well against a variety of alternative distributions (Horswell and Looney (24), Mardia and Foster (25)). Since these tests are based on the same principles as the one above, we expect them to behave similar to that of (7).

IV. THE MONTE CARLO EXPERIMENT

The design of a good Monte Carlo study is dependent on (a) what factors are expected to affect the properties of the test under investigation and (b) what criteria are being used to judge the results. We will in what follows look at these questions in more details.

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 (i.e., the probability of rejecting the null when true) is close to the nominal size (used to calculate the critical values). Given that the actual size is a reasonable approximation to the nominal size, we then wish to investigate the actual power of the test (i.e., the probability of rejecting the null when false) for a number of different alternative hypotheses.

First, we want to study the size property using the asymptotic distribution of the statistic. Second, we wish to use the fact that the statistic is invariant to linear transformations in order to generate empirical critical values, and study the size and power properties of this approach. Several factors are expected to affect the properties of the JM test. We will here try to cover various combinations of some of these in order to examine the properties of the test. In Tables I and II, we

present a summary of the Monte Carlo design used in this paper. Relevant factors considered in this study are

i.   The nominal size, α.

ii.   The number of equations P.

iii.   The sample size n.

iv.   The alternative distribution H_A.

A number of other factors can also affect the properties of the JM test. The impact of the biases in the estimated residual moments due to the $\mathbf{M}$ matrix on the JM test is unknown. The distribution of $\mathbf{X}$ and the stochastic properties of the residual are thus obvious candidates to examine. In a later section of this study, we will consider these in some more detail. Another relevant feature is to examine the robustness of the JM test against autocorrelation (which is frequently appearing in economic data). This will as well be treated in the experiment.

Our primary interest lies in analysis of system wise tests, and thus the number of equations to be estimated is of central importance. The number of equations in econometrics is rarely larger than 10. Based on this, we examine k = 1, 2, 5, 10. Since the test is known to be consistent against any distribution with non-normal skewness or kurtosis, and the residuals are also known to be consistent estimates of the disturbances, it follows that the whole test is consistent. We will therefore focus on small samples. As we are also interested in the interaction between sample size and number of equations, the number of degrees of freedom (v) is held constant when comparing models with different numbers of equations. As previously mentioned, one of the objective of this study is to investigate the properties of the JM test in small samples, hence we used values of v ranging from 5 to 125 degrees of freedom.

Another purpose of this study is to examine how fast the size of the test converges to the actual size. Since the experiment is performed using a finite number of replicates, we must be able to distinguish simulation fluctuations from biases in the test. One possibility to do this is to calculate an approximate 99% confidence interval for the actual size α:

$$\hat{\alpha} \pm 2.575 \sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{R}}$$

where $\hat{\alpha}$ is the estimated size and R is the number of replicates. To judge the reasonability of the results, we require that the estimated size should lie within the 99% confidence interval of the actual size. For example, if we consider a nominal size of 5%, and when we operate 100 000 replications, we define a result as reasonable if the estimated size lies between 0,0482 and 0,0518. Even if the actual size of a test correctly corresponds to the nominal size, the test will be of little use if it does not have sufficient power to reject a false null hypothesis. In the rest of this section we will consider this question in some more detail. There are two different ways that a distribution can depart from multivariate normal kurtosis:

i.   At least one marginal kurtosis, $\gamma_{a,b,c,d}$ for $a = b = c = d$, is different from 3.

ii.   Other non-univariate fourth-order moments have non-MVN values.

Horswell and Looney (24) refer to these departures as "visible" and "invisible" kurtosis respectively. In a similar way, the departures from MVN can also be visible or non-visible skewness, as well as combinations of the both. In order to test for departures of the "invisible" type, we make use of the non-MVN distribution of Khintchine. Let $X_i = \pi_i R_i U$, $i = 1,2,...P$, where

$$U \sim U[0,1], \quad R_i \sim [\Gamma(\lambda,p)]^\tau \quad \text{and} \quad p(\pi_i) = \begin{cases} 0.5 & \pi_i = 1 \\ 0.5 & \pi_i = -1 \end{cases}.$$

The coefficient of kurtosis is then determined by $\beta_2 = \dfrac{\Gamma(p+4\tau)\Gamma(p)}{\Gamma(p+2\tau)^2}$ (Johnson (26)). Fixating $\beta_2 = 3$ and choosing an arbitrary value for τ, we can obtain the value of p by numerical optimisation. In this study we choose $\tau_1 = 0.1$, $\tau_2 = 3.5$ with the corresponding values $p_1 = 0.12757$ $p_2 = 89.507$.

The two Khintchine variables will be denoted as K_1 and K_2 respectively. In addition, we will examine two cases: $R_1 = R_2 = ... = R_p$ and $R_i \neq R_j$ respectively. In order to get an idea of the shape of these (marginal) distributions K_1 and K_2, 500 000 pseudo observations have been simulated for each of them. In addition, K_1 and K_2 have the same scale so that $\mu = 0$, $\mu_2 = 1$, $\mu_3 = 0$, $\mu_4 = 3$ for both of them. Their densities are displayed below.
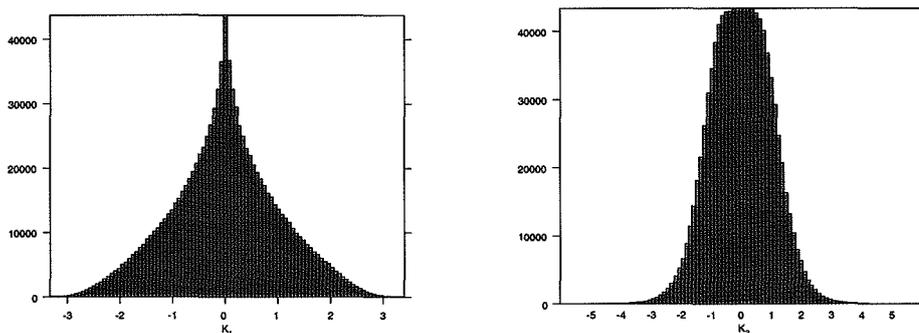
Note that the K_2 distribution would be almost impossible to distinguish from the normal distribution by simply studying the histogram. Moreover, since it is well known that the disturbances in economic data tend to have "heavy tails", we also use marginal t-distributions as alternative distributions. Another purpose of this paper is to investigate if the properties of the JM test will be adversely affected by an AR(1) or an MA(1) structure in the error terms. These two processes will so be included in the experiment.

TABLE I.
Values of Factors Held Constant that May Affect the JM Tests

Factor	Value
Properties of X in repeated samples	Stochastic
Structure of the error terms	White noise, AR and MA
Number of X variables	5
Mean of X variables	0
Order of error AR processes	1
Order of error MA processes	1

TABLE II.

Values of Factors that Vary for Different Models-Size and Power Calculations

Factor	Symbol	Design
Number of equations	n	1, 2, 5, 10
Degrees of freedom	ν	5, 15, . . ., 125
Nominal size	α	1%, 5%
AR parameter for errors	ϕ	0, .3, .5, .7, .95
MA parameter for errors	θ	0, .3, .5, .7, .95
Distribution of X variables		Normal, $t_{(7)}$, $t_{(5)}$, $t_{(3)}$, $t_{(1)}$
Distribution of error terms		Normal
(only for power calculations)		$t_{(7)}$, $t_{(5)}$, $t_{(3)}$, $t_{(1)}$, and K_1, K_2

V. RESULTS

In this section, we present the most interesting results along with results of the main dominating effects of our Monte Carlo experiment regarding both size and power properties of the JM test. Since the experiment is quite extensive this must be done in a fairly compact manner, full results are, however, available from the authors.

SIZE PROPERTIES:

In this subsection, results concerning the size properties of the JM test are presented in graphic forms. These plots make it possible and easy to find out situations under which the tests may systematically over- or under-reject, or reject the null hypothesis about the right proportion of the time. The first four graphs show the empirical size of the JM test under ideal premises (i.e., iid errors), while graph 5-6 concerns the empirical size when the independency assumption is violated. The size has been estimated from 1 million Monte Carlo replicates.

Figure 1. The estimated size for the JM test at 1% and 5% levels using asymptotic null distribution and N(0,1) regressors. The upper line corresponds to the 5% level.

P = 1:   P = 2:

P = 5:   P = 10:
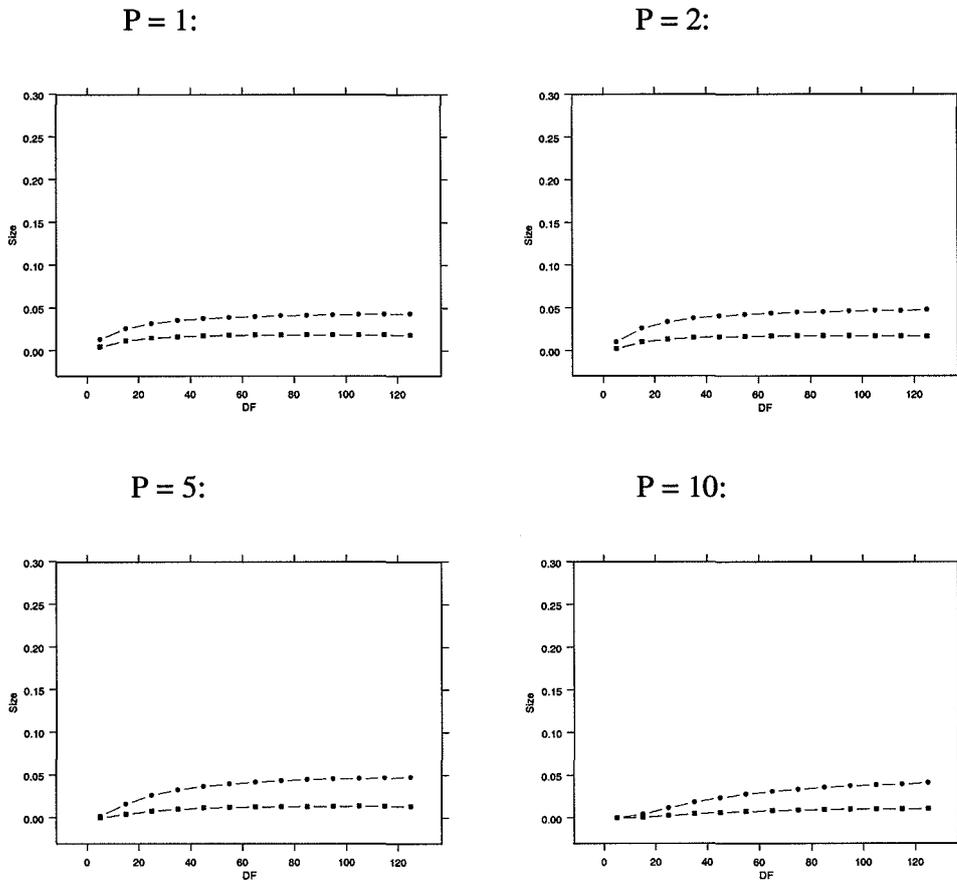
In Figure 1, we present the results of the estimated JM test at the 5% and 1% nominal sizes in systems ranging from one to ten equations where the regressors follow a N(0,1) distribution. Looking at these graphs for the 5% nominal size, we can see that the test does not perform very well, in the sense that it under rejects, especially in small samples and large systems. On the contrary, when looking at the 1% nominal size, the test tends to over reject. One possible explanation may be that the small sample distribution of D_p is skewed, relative to the chi-square distribution. We will, however, not pay any further attention on this problem as the empirical critical values are easy to obtain, and performs well (see Fig 2).

Figure 2. The estimated size for the JM test at 5% and 1% levels using empirical critical values (simulated with 10 million replicates) and N(0,1) regressors. The upper line corresponds to 5% level.
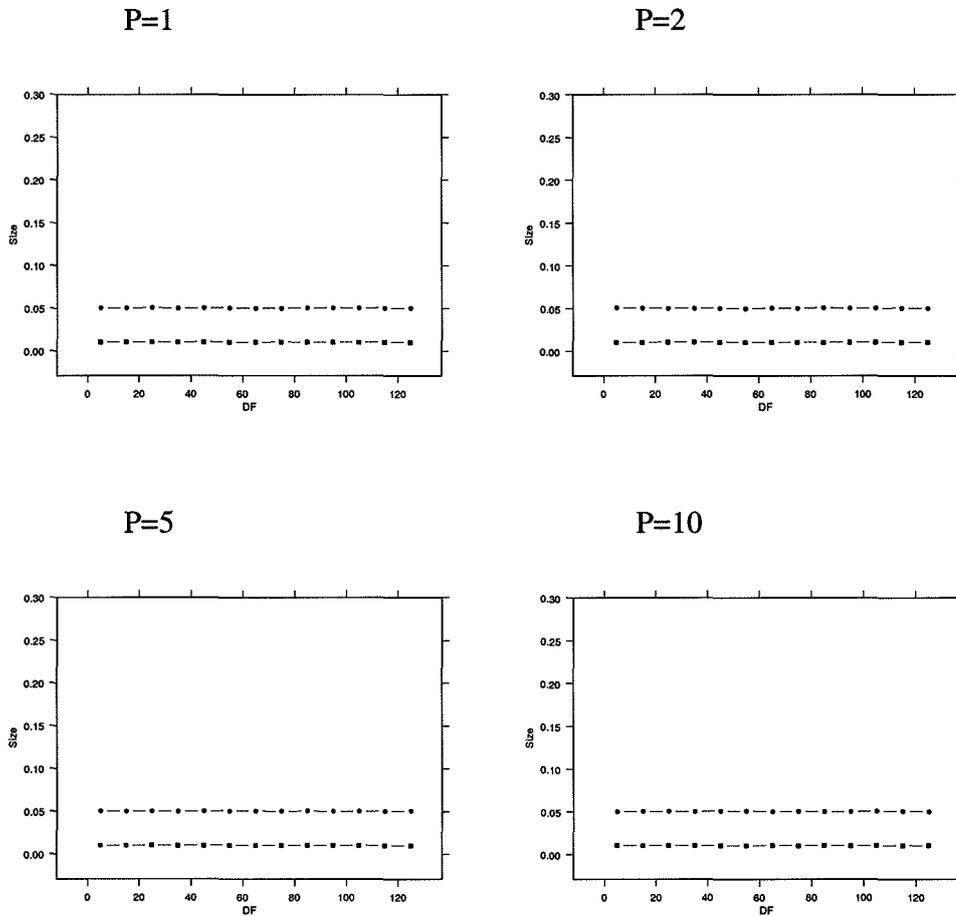
In Figure 2 we clearly see that empirical critical values are indeed a good tool for controlling the size of the JM test. Even though the empirical critical values have been generated from a model with just an intercept, the size level is maintained when using a regressor matrix of 5 normally distributed regressors. The test performs satisfactorily even in small samples and large systems of equations. Also, the fact that the empirical critical values have been calculated from a finite number of observations seems to be negligible as the random fluctuations are of smaller magnitude than the third digit.

Figure 3. The estimated size for the JM test at 5% level using empirical critical values and $t_{(1)}$, $t_{(3)}$, $t_{(5)}$, $t_{(7)}$ distributed regressors. The upper line corresponds to $t_{(1)}$, the lowest to $t_{(7)}$.

P=1

P=2

P=5

P=10
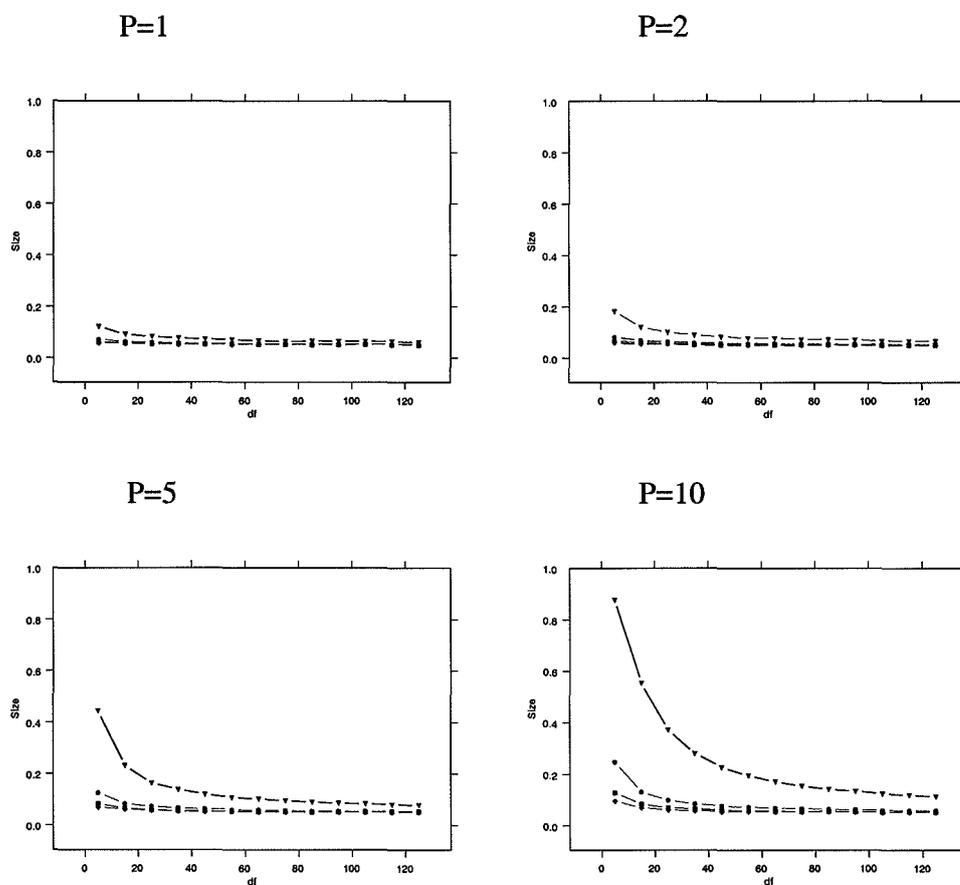
In Figure 3, we present the results of the estimated size of the JM test, at the 5% nominal size in systems ranging from one to ten equations where the regressors follow t-distribution with different degrees of freedom. The same set of empirical critical values have been used as in Figure 2. When considering Figure 3, we can see the impact of heavy tailed regressors on the JM test. The test performs well in small systems of equations or in large samples, while it performs extremely badly, in the sense that it over rejects, in small samples and large systems.

Figure 4. The estimated size for the JM test at the 5% level using empirical critical values and $t_{(1)}$ distributed regressors and homogenized OLS residuals.

P=1

P=2

P=5

P=10
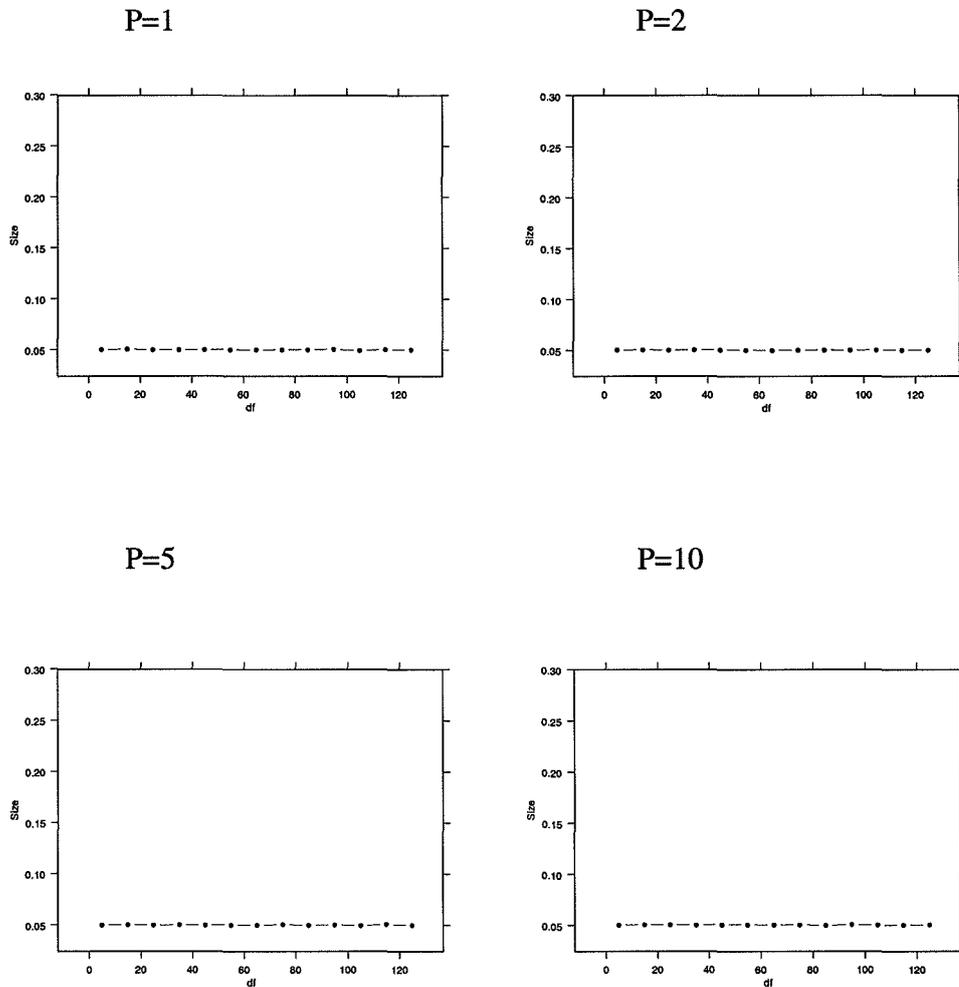
Figure 4 visualises that homogenised regressors is indeed a good remedy for the bad effect of the fat tailed distributed regressors (which we previously mentioned in Section II). Again, the same set of empirical critical values as in Figure 2 have been used.

Figure 5. *The estimated size for the JM test at 5% level using empirical critical values and* AR(1), $\phi = 0.95$, 0.7, 0.5, 0.3 *distributed disturbances and* N(0,1)*distributed regressors. The upper line corresponds to* $\phi = 0.95$, *while the lowermost to* $\phi = 0.3$.

P = 1   P = 2

P = 5   P = 10
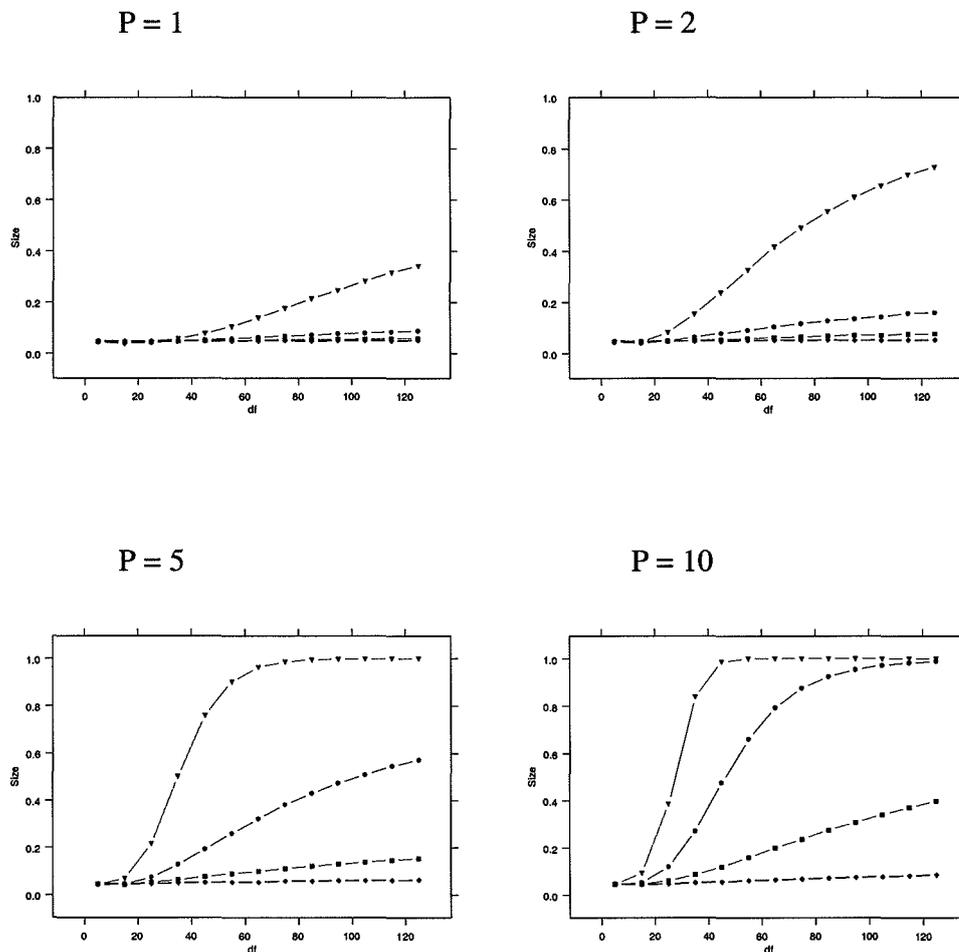
Figure 5 reveal that the JM test is sensitive to autocorrelation. The AR(1) process seems to have a serious effect on the properties of the JM test. For example, if we are analysing a system of 5 equations with 80 observations when the disturbances follow an AR(1) process with intensity parameter 0.95 (which is not an unrealistic case), we will reject the null hypothesis (when the null is true) in 99% of the cases in repeated sampling!

Figure 6. The estimated size for the JM test at 5% level using empirical critical values and MA(1), $\theta = 0.95$, 0.7, 0.5, 0.3 *distributed disturbances and* N$(0,1)$*distributed regressors. The uppermost lines corresponds to* $\theta = 0.9$, *while the lowermost line corresponds to* $\theta = 0.3$.

P = 1   P = 2

P = 5   P = 10
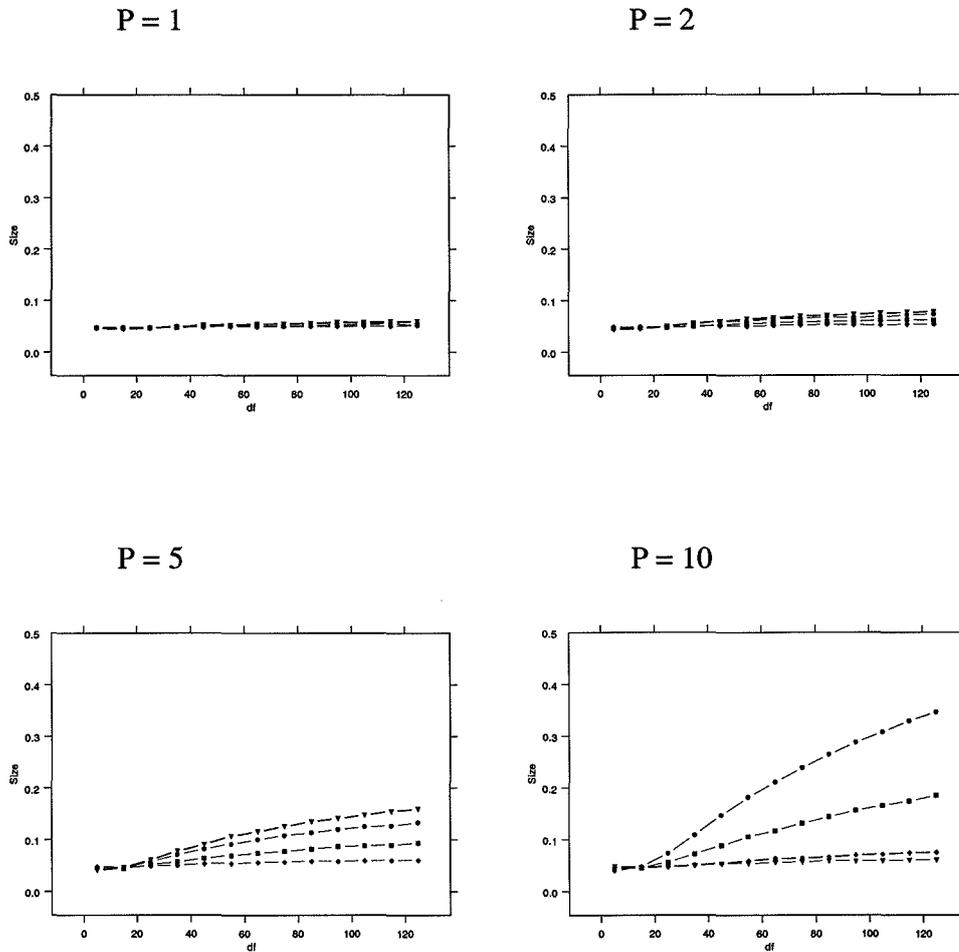
In Figure 6, we present the results of the JM test when the error terms follow an MA(1) structure with different parameters. The effect of the MA(1) process on the properties of the JM test is less than that of the AR(1) process, but it is still serious especially in large systems of equations. For the special case of P = 1 the test remain robust though, at least within the examined range of degrees of freedom.

POWER PROPERTIES:

In this subsection, results concerning the power properties of the JM test against two families of alternative distributions are presented. All the presented results here are at the 5% significance level, using empirical critical values. The power functions of the JM test were estimated by calculating rejection frequencies from 100 000 replications for error terms that follow $t_{(7)}$, $t_{(5)}$, $t_{(3)}$, $t_{(1)}$, and K_1, K_2 distributions.

Figure 7. The estimated power for the JM test at 5% level using empirical critical values with $t_{(1)}$, $t_{(3)}$, $t_{(5)}$, $t_{(7)}$ distributed regressors. The uppermost lines correspond to $t(1)$, while the lowermost lines correspond to $t(7)$.
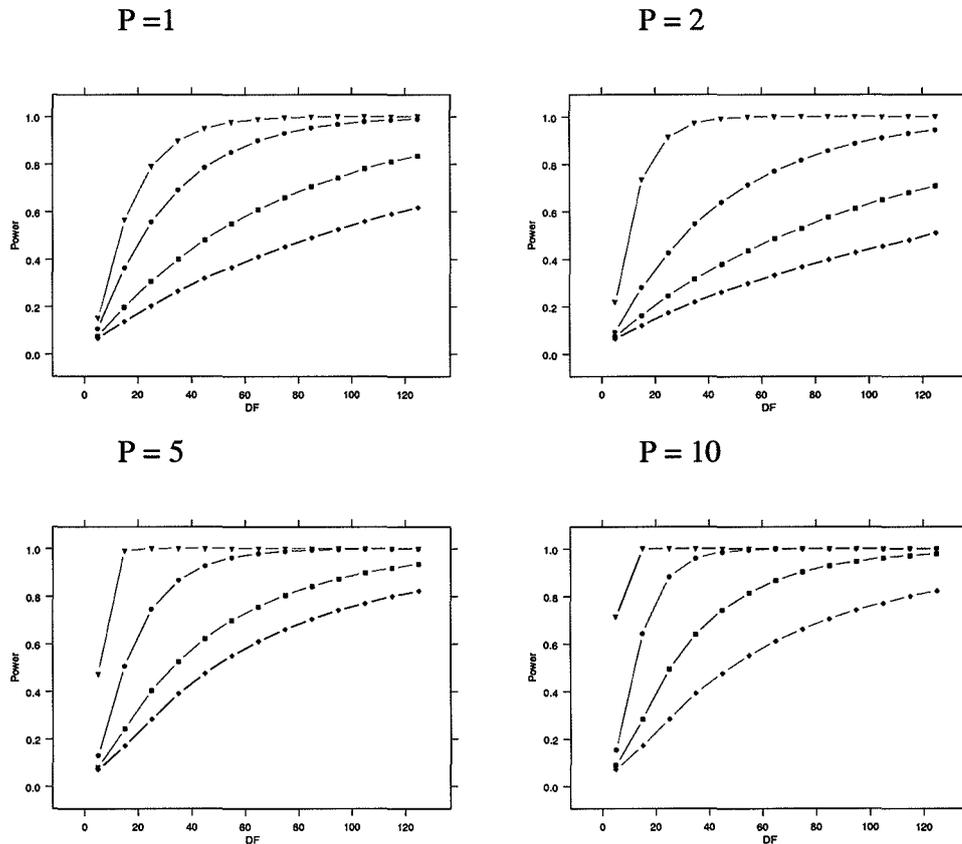
In the figure above we observe high power against t-distribution with few df. In fact, the kurtosis for $t_{(1)}$, $t_{(3)}$ does not exist at all. Still, the JM test seems to detect these distributions perfectly.

Figure 8. The estimated power for the JM test using empirical critical values and K_1, K_2 distributed disturbances. The two uppermost lines correspond to K_1 and K_2 with identical uniform components, while the lower lines correspond to K_1 with independent uniform components.

P=1   P=2

P=5   P=10
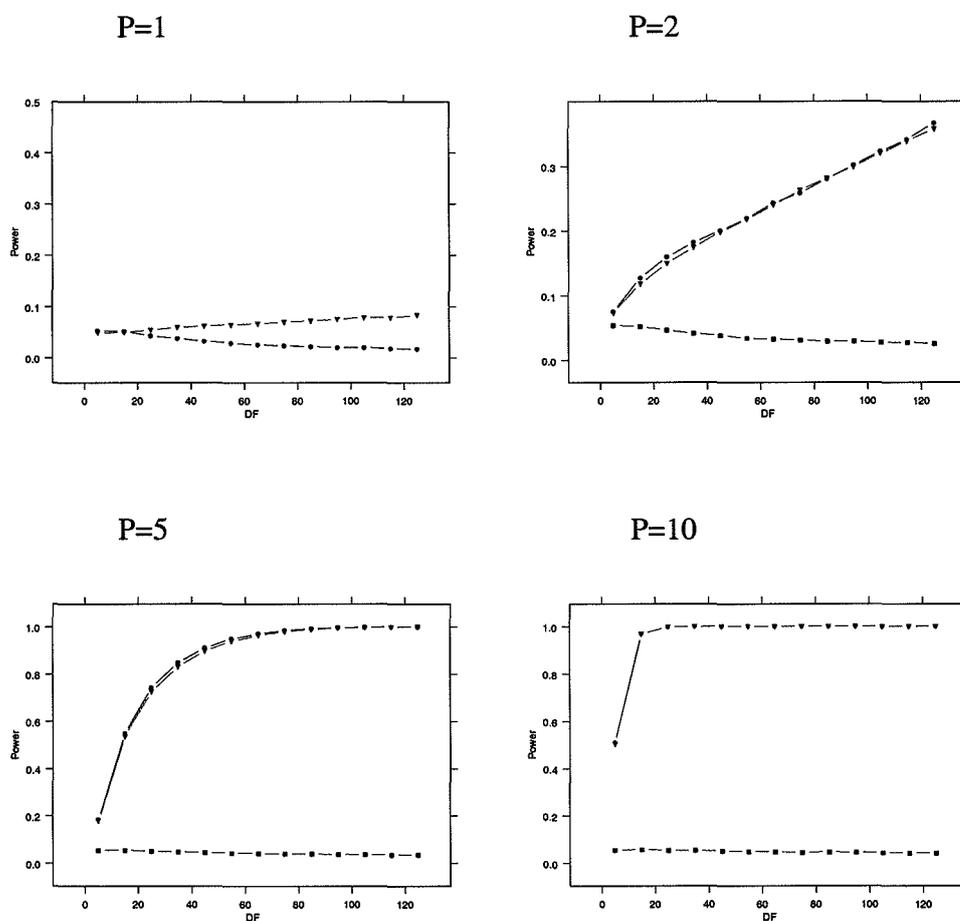
Figure 8 reveals some interesting feature. The power against the K_1 distribution with independent gamma generators is literally zero. In fact, the power even seems to be lower than the nominal size, which may seem curious. On the contrary, the power against the K_1 and K_2 with identical gamma generators is very high. In addition, the test appears to be invariant to the value of τ.

VI. SUMMARY AND CONCLUTIONS

In this paper we have studied the properties of system-wise JM test for non-normality when the error terms follow a normal distribution, t-distribution with different degrees of freedom, and non-MVN distribution of Khintchine.

The investigation has been carried out using Monte Carlo simulations. Several models were investigated regarding the size of the tests, where the number of equations, degrees of freedom and stochastic properties of the exogenous variables have been varied. For each model we have performed 1000,000 replications and studied two different nominal sizes. The power properties have been investigated for using 100,000 replications per model, where in addition to the properties mentioned above the distribution of the error terms have also been varied.

Since it is well known that both of the components of the test statistic D_p converge slowly to their asymptotic distribution, we expect the D_p statistic to converge slowly as well. This fact is clearly reflected in the experiment. What may seem surprising is that the size is overestimated at the 1% level while it is underestimated at the 5% level. One possible explanation may be that the small sample distribution of D_p is skewed, relative to the chi-square distribution. When using the empirical critical values instead, the test has shown to perform as expected. Consequently, we recommend that empirical critical values should always be used for the JM test.

The effect of the heavy-tailed or extremely skewed regressors has shown to be substantial, especially in small samples and large systems. However, the homogenised OLS residuals have indeed shown to be a good remedy for this problem.

A much more disturbing fact is that the impact of the autocorrelated disturbances on the JM test is devastating. For high autocorrelation parameter and large systems, the JM test tends to reject 100% of the time under the null hypothesis. In fact, the test is not consistent when the auto covariance is non-zero

for any lag. This may be a serious problem in using moment-based tests for non-normality.

The power of the JM test seems to be high against most of the treated alternative distributions. In general, the power increases with the number of equations. The marginal distributions that have non-normal fourth moment, results in rather high power, even for relatively normal-close distributions as $t_{(7)}$.

In stark contrast, the power against the K_1 distribution with independent gamma generators is literally zero. In fact, the power even seems to be lower than the nominal size, which may seam curious. On the contrary, the power against the K_1 and K_2 distributions with independent gamma generators is very high. This is an illuminating result, since the power following from performing equation-wise tests would be zero. Another interesting feature is that the powers for K_1 and K_2 are identical. This suggests that the test is invariant to the value of τ.

One obvious weakness of the JM test is that it is non-diagnostic. A natural question is what to do if a diagnostic test is needed. One possible solution is to try to find sufficient conditions, if possible, for the sample kurtosis and skewness to be strictly diagnostic. However, the assumption of symmetric distribution is quite reasonable, since our variable of interest is indeed noise, and noise should have the property $f(\varepsilon) = f(-\varepsilon)$. Still, in order to have power against skewed noise due to misspecification, we feel that an omnibus test should be used, rather than relying totally on the symmetry assumption. In fact, deviation from normality, as well as autocorrelation, can be viewed as misspecification of the model.

One important issue that is of great relevance is in what situations non-normality is so serious that it ruins the whole modelling procedure, i.e. when should normal-theory be abandoned (in favour for e.g. non-parametrical methods)? This question is however beyond the scope of this paper, but we would like to stress that this issue is important to consider when judging the result in this paper (or performing non-normality test on real data).

References

(1) Geary, R. C. Testing for Normality. Biometrika, **1947**, *34*, 209-242.

(2) Kolmogorov, A. N. Sulla determinazione empirica di une legge di distribuzione. G. Ist. Att. uari, **1933**, 83-91.

(3) Pierce, D. A.; Kopecky, K. J. Testing goodness of fit for the distribution of errors in regression models. Biometrika, **1979**, *66* (1), 1-5.

(4) Cox, D. R.; Hinkley. D. V. Theoretical Statistics. 2 ed. Chapman and Hall Ltd, **1974**.

(5) Theil, H. The Analysis of Disturbances in Regression Analysis. Journal of the American Statistical Association, **1965**, *60*, 1067-79.

(6) Hedayat, A.; Robson, D. S. Independent Stepwise Residuals for Testing for Homoscedastsity. Journal of the American Statistical Association, **1970**, *65*, 1573-1581.

(7) Brown, R. L.; Durbin, J.; Evans, J. M. Techniques for Testing the Constancy of Regression Relationships over Time. Journal of the Royal Statistical Society, Series B, **1975**, *37*, 149-172.

(8) Bryc, W. The Normal Distribution Characterizations with Applications. Lecture Notes in Statistics, Springer Verlag, **1995**; Vol. 100.

(9) Lucaks, E.; Laha, R. G. Applications of Characteristic Functions. 1 ed. Griffin´s Statistical Monographs & Courses, **1963**.

(10) Vasicek, O. A test for normality based on sample entropy. Journal of the Royal Statistical Society. B, **1976**, *38*, 54-59.

(11) Rao, C. R. *Linear Statistical Inference.* 2 ed. Wiley, **1973**.

(12) Epps, T. W. Characteristic Functions and Their Empirical Counterparts - Geometrical Interpretations and Applications to Statistical Inference. American Statistician, **1993**, *47* (1), 33-38.

(13) Heyde, C. C. On a Property of the Lognormal Distribution. Journal of the Royal Statistical Society. B, **1963**, *25*, 392-393.

(14) Kendall, M.; Stuart, A. *The Advanced Theory of Statistic, Volume 1, Distribution Theory.* 4 ed., Charles Griffin&Company Limited, **1976**.

(15) Rayner, J. C. W.; Best, D. J.; Mathews, K. L. Interpreting the skewness coeficient. Communications in statistics-theory and methods, **1995**, *24* (3), 593-600.

(16) Horswell, R. L.; Looney, S. W. Diagnostic Limitations of Skewness Coefficients in assessing Departures from Univariate and Multivariate Normality. Communications in statistics-simulation and computation, **1993**, *22* (2), 437-459.

(17) Churchhill, E. Information Given by Odd Moments. Annals of Mathematical Statistics. **1946**, *17*, 244-246.

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

(19) Mardia, K. V. Meassures of Multivariate skewness and Kurtosis with Applications. Biometrica, **1970**, *57* (3), 519-530.

(20) Malkovich, J. F.; Afifi, A. A. On Tests For Multivariate Normality. Journal of the American Statistical Association, **1973**, *68*, 176-179.

(21) Srivastava, M. S. A Meassure of Skewness and Normality and a Graphical Method for Assessing Multivariate Normality. Statistics&Probability Letters, **1984**, *2*, 263-267.

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

(23) Huang, C. J.; Bolch, B. W. On the Testing of Regression Disturbances for Normality. Journal of the American Statistical Association, **1974**, *69* (346), 330-335.

(24) Horswell, R. L.; Looney, S. W. A Comparison of Tests for Multivariate Normality that are Based on Meassures of Multivariate Skewness and Kurtosis. Journal of Statistical Computation and Simulation, **1992**, *42*, 21-38.

(25) Mardia, K. V.; Foster, K. Omibus Tests of Multinormality Based on Skewness and kurtosis. Communications in Statistics-theory and methods. **1983**, *12* (2), 207-221.

(26) Johnson, M. E. *Multivariate Statistical Simulation.* 1 ed. Wiley, **1987**.

Testing for Non-normality in Multivariate Regression
with Nonspherical Disturbances

H.E.T. Holgersson

Department of Statistics

Gothenburg University

P.O. Box 660

SE-405 30 Gothenburg, Sweden

ABSTRACT

The problem of testing for non-normality in multivariate regression models when nonspherical disturbances are present is considered by means of Monte Carlo experiments. We give a reason why moment based non-normality tests, e.g. the Jarque&Bera test and multivariate extensions, generally fail when the data is not independent identically distributed. We propose several possible choices of proxy variables to the unobservable errors, which are applicable to non-normality testing as long as the structure of the covariance matrix is known. However, simulations reveal that even a small misspecification of the covariance structure may well lead to an inconsistent test procedure in the sense that the size will limit unity. We argue that the use of regular non-normality tests on variables with a complicated data generating process, such as in economic applications, is dubious.

I. Introduction

A frequently occurring problem in statistics is that many diagnostic tests are not unique, in the sense that tests to one certain specification may be highly sensitive to another specification. Thus, the practitioner may be totally misled, trying to re-specify his/her model in a wrong direction. One specific example is given in Holgersson and Shukur (1), 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. Indeed, the non-normality test appears to be consistent against autocorrelation. Also, autocorrelation or heteroscedasticity is well known to have the effect of making point estimates of regression parameters inefficient. A natural approach is then to handle the autocorrelation/heteroscedasticity first and then apply the non-normality test, so that the practitioner may test that the final model is not only iid, but also iid normal. We will, in what follows, consider this approach in some more detail.

The main purpose of this paper is to investigate the usefulness of various residuals resulting from a generalised least square (GLS) or feasible generalised least square (FGLS) estimation to test for non-normality, including the possibility of a small (and thus realistic) misspecification of the covariance matrix.

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 (2) test, will not converge to its asymptotic null distribution when the variable is non-iid. Section **IV** is concerned with possible choices of observable proxy variables to the unobservable disturbances along with some asymptotic properties. In section **V** we present the Jarque and Mckenzie (3) test and show how an exact size can be obtained if the covariance matrix is known. 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**.

1

II. Model specification

The model considered in this paper is the multivariate regression model

$$\mathbf{Y}_{(N \times P)} = \mathbf{X}_{(N \times k)} \mathbf{B}_{(k \times P)} + \boldsymbol{\varepsilon}_{(N \times P)},$$ (1.1)

where, by assumption, $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2 \Omega$, Ω a positive definite (P.D.) matrix, $E\left[\mathbf{X}'\Omega^{-1}\boldsymbol{\varepsilon}\right] = \mathbf{0}$. Further on, define $\tilde{\mathbf{X}} := \Omega^{-1/2}\mathbf{X}$ and let $d_{Nk}^2 := \sum_{i=1}^{N} \tilde{x}_{il}^2$, $l = 1, 2, ..., k$. We then assume that $plim\, d_{Nk}^2 = \infty$ and that $plim\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}$ is P.D. A frequently made assumption of (1.1) is that $\boldsymbol{\varepsilon}$ is normally distributed. That a variable is normally distributed is a somewhat vague property that 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 not the case as most standard tests are under the null hypothesis assuming that $\boldsymbol{\varepsilon}$ is (multivariate) *iid* normal. Thus we invoke *two* assumptions in the null hypothesis, which is often overlooked. The density function of a multivariate normally distributed variable is given by

$$f(\boldsymbol{\varepsilon}') = \left[(2\pi)^2 |\Sigma|\right]^{-\frac{N}{2}} |\Xi|^{-\frac{p}{2}} exp\left[tr\left(-\frac{1}{2}\Sigma^{-1}\boldsymbol{\varepsilon}\Xi^{-1}\boldsymbol{\varepsilon}'\right)\right], \text{ or } \boldsymbol{\varepsilon}' \sim N_{p,N}\left(\mathbf{0}, \Sigma_{(p \times p)}, \Xi_{(N \times N)}\right).$$

Whenever $\Xi \neq \sigma^2 \mathbf{I}$, the errors are said to be *nonspherical*. Two such cases which will be considered in detail are heteroscedasticity and autocorrelation. Apart from difficulty with assessing normality (see Section III), the main problem with non-spherical errors is that the point estimates of the regression parameters will be inefficient, and that the interval estimates can be underestimated. If the variables are not iid due to autocorrelation, the problem is usually overcome by re-specifying the model. However this approach often fails, due to the complexity of the data generating process. Heteroscedasticity often arises naturally in a model due to the fact that the variance is a function of the regressors. It is well known for example, that the variability in savings is larger for those with large income than for those with a small income. In such situations the GLS method is commonly used. For multivariate regression we can describe the method as follows:

2

Let $E[\varepsilon\varepsilon'] \equiv \Xi \equiv \sigma^2\Omega$. If Ω is known, we can perform the transformation

$$\mathbf{Y} = \mathbf{XB} + \varepsilon \rightarrow \Omega^{-1/2}\mathbf{Y} = \Omega^{-1/2}\mathbf{X} + \Omega^{-1/2}\varepsilon, \text{ so that } E\left[\Omega^{-1/2}\varepsilon\left(\Omega^{-1/2}\varepsilon\right)'\right] = \sigma^2\mathbf{I}. \text{ The OLS}$$

estimates of the regression parameters of the transformed model above are both unbiased and efficient. Additionally, the transformed errors $\Omega^{-1/2}\varepsilon$ are useful in order to make inferences of the true errors, e.g. to test for non-normality as in our case, since the covariance is scalar. Unfortunately, Ω is usually unknown, meaning that we have to consider an estimated version and rely on asymptotic properties. Furthermore, ε is unobservable, which means that it has to be estimated first, resulting in a three-stage estimation procedure. It is therefore relevant to examine the small sample properties of this approach. We will discuss these problems further in chapter **IV**.

III.   Failure of skewness and kurtosis testing of non-iid variables.

The skewness and kurtosis coefficients are frequently used to test for non-normality. Particularly popular is the Jarque and Bera (2) omnibus test (JB), with its statistic defined as $\left\{N\hat{\gamma}^2_1/6\right\} + \left\{N\left(\hat{\gamma}_2 - 3\right)^2/24\right\}$, where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the sample skewness and kurtosis coefficients. When the target variable is iid normally distributed, the standardised statistics above are each asymptotically $\chi^2_{(1)}$ distributed. If the observations are not independent, i.e. if $\sigma^2\Omega \neq \sigma^2\mathbf{I}$ due to autocorrelation, the JB statistic will not converge to its null distribution, as is exemplified in Holgersson and Shukur (1). To obtain an explicit reason for the non-convergence of the statistic in this case, we use the result of Lomnicki (4): Consider a univariate time series

$X_t = \sum_{i=0}^{\infty} h_i \varepsilon_{t-i}$, where $\sum_{i=1}^{\infty}|h_i| < \infty$ and $\varepsilon_t \sim iid\ N\left(0, \sigma^2\right)$ with autocorrelations

$\rho_j := E\left(X_t X_{t-j}\right)/E\left(X_t^2\right)$. Lomnicki showed that, for a one-dimensional random

variable, $\left\{\sqrt{N\hat{\gamma}_{1,1}/6}\right\} \xrightarrow{\ell} N\left(0, \sum_{k=-\infty}^{k=\infty}\rho_k^3\right)$ and $\left\{\sqrt{N}\left(\hat{\gamma}_{2,1} - 3\right)/\sqrt{24}\right\} \xrightarrow{\ell} N\left(0, \sum_{k=-\infty}^{k=\infty}\rho_k^4\right)$.

From that follows immediately that

$$\rho_k \neq 0 \Rightarrow \left\{N\hat{\gamma}_{1,1}/6\right\} + \left\{N\left(\hat{\gamma}_{2,1} - 3\right)^2/24\right\} \xcancel{\xrightarrow{\ell}} \chi^2_{(2)} \quad \forall_{k \neq 0}. \tag{3.1}$$

To examine the null distribution of the JB statistic for the case where $\sigma^2 \Omega \neq \sigma^2 I$ due to heteroscedasticity rather than autocorrelation, we use the result of Kendall and Stuart (5), that, given normality:

$$E\left[\left(\frac{m_3}{m_2^{3/2}}\right)^k\right] = \frac{E\left[m_3^k\right]}{E\left[m_2^{3k/2}\right]}, \quad \text{and} \quad E\left[\left(\frac{m_4}{m_2^2}\right)^k\right] = \frac{E\left[m_4^k\right]}{E\left[m_2^{k2}\right]}.$$

The moments of the univariate skewness and kurtosis coefficients for heterogeneous observations can then be obtained by assuming that $\{x_i\}_{i=1}^n$ are independently distributed $N(0,\sigma_i^2)$, and that $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^n \sigma_i^8 = 0$. The numerators of the quotients above are given in (5), from where we have (using leading term approximations):

$$E\left[m_3^2\right] = \frac{1}{n^2}E\left[\sum_{i=1}^n x_i^6 + (9/n^2)\left(\sum_{j\neq k\neq l} x_j^2 x_k^2 x_l^2\right) - (6/n)\left(\sum_{j\neq k} x_j^4 x_k^2\right)\right] + o\left(n^{-1/2}\right) =$$

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

$$E[m_4] = E\left[(1-(4/n))\sum_{j=1}^n x_j^4\right]/n = 3\sum_{j=1}^n \sigma_j^4 / n + o\left(n^{-1/2}\right) \quad (3.3)$$

Following Magnus (6) we have (again using leading term approximations):

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

$$E\left[m_2^3\right] = \left[(tr(\Omega))^3 + 6tr(\Omega)tr(\Omega^2) + 8tr(\Omega^3)\right]/n^3 = \sum_{j\neq k\neq l} \sigma_j^2 \sigma_k^2 \sigma_l^2 / n^3 + o\left(n^{-1/2}\right) \quad (3.5)$$

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

$$\left(\sigma_1^2 \neq \sigma_2^2 \neq ... \neq \sigma_n^2\right) \Rightarrow \left\{N\hat{\gamma}_{1,1}/6\right\} + \left\{N\left(\hat{\gamma}_{2,1}-3\right)^2/24\right\} \xrightarrow{\ell} \chi_{(2)}^2. \quad (3.6)$$

4

In addition, measures of multivariate skewness and kurtosis that are extensions of those above, e.g. Lütkepohl and Theilen (7), Malkovich and Afifi (8) or Jarque&McKenzie (see Section V) will in general not converge to its null distribution when the covariance of the disturbance vector is non-scalar. 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 iid, as they usually measure the complete distribution of the target variable. Consequently, accepted null hypothesis will indicate that the data is normally distributed (though not necessarily iid normally distributed, as consistency of those tests against heteroscedasticity or autocorrelation remains to be shown). However, a rejected null hypothesis may either be due to non-normality or non-scalar covariance matrix. Thus standard tests for non-normality are diagnostic only if it is known that the target variable is (at least asymptotically) iid. As non-iid data are frequently occurring, the impact on these for multivariate non-normality tests is of great interest.

Another issue is how to overcome the problem. There are two obvious possibilities. One is to construct tests that are robust to non-iid data (i.e. to construct tests that assess the distribution of ε). The second, which we will consider here, is rather to assess the distribution of transformed variables with scalar covariance matrixes, e.g. $\Omega^{-1/2}\varepsilon$. Lütkepohl and Schneider (9) examined the possibility of using residuals for evaluation of the distribution of the random component for the case of pure autoregressive processes. In our case with exogenous information, the most natural approach is to use GLS/FGLS estimation, thus covering heteroscedasticity as well, though residuals from autoregressive processes can also be used. In fact, there exist many possible choices of variables that are asymptotically iid, and therefore may be used for non-normality testing. For example, one may use residuals from FGLS residuals, residuals from iterative FGLS or transformed ordinary least square residuals. In this paper we will give a few examples of such useful variables, apply a non-normality test on them and compare the size, power and robustness. In the next section we present the variables used in the study, along with some asymptotic properties.

IV.   Proxy variables to the unobservable disturbances.

Often we want to make some inferences on the disturbance component ε of (1.1). As this is unobservable we will have to use an observable proxy variable instead, usually the residuals $\hat{\varepsilon} = Y - \hat{Y}$. A specifically useful property of these that ensures their usefulness in diagnostic testing is that they converge in probability to the true disturbances, i.e. $\left|\hat{\varepsilon}_{i,n} - \varepsilon_i\right| \overset{p}{\to} 0$. Indeed, it is also well known that $\hat{\varepsilon}$ is the best linear unbiased estimate to ε (the word "estimate" applied to a random variable is sometimes subject to debate, the alternative notation being "prediction", but we prefer "estimate" in order to stress that we are not interested in guessing future values). However, our problem is to estimate $\Omega^{-1/2}\varepsilon$ which is a somewhat different task, and we will see below that these estimates do not in general converge in probability to their unobservable counterpart. In fact, none of the estimates of the unobservable variable that we will consider here are residuals in the sense that they are defined as $\hat{\varepsilon} = Y - \hat{Y}$, but they are residuals in the sense that they will limit the distribution of the disturbances. Rather, they can be thought of as quasi residuals, although we will refer to them as residuals anyhow, in lack of a better word. Below we will discuss different possibilities of choosing such residuals. As the residual vector cannot formally have a limit (because the dimension grows with n) we will state the following:

Definition:

i. $X_n \overset{p\bullet}{\to} \theta_n := X_{i,n} \overset{p}{\to} \theta_i$, (i = 1, 2, ...,n), i.e. elementwise convergence in probability, though the dimension of θ is allowed to grow with n.

ii. $X_n \overset{\ell\bullet}{\to} \delta_n := X_{i,n} \overset{\ell}{\to} \delta_i$, (i = 1, 2, ...,n), i.e. elementwise convergence in law, though the dimension of δ is allowed to grow with n.

This definition will make it possible for us to adequately talk about limiting properties of residual vectors in a meaningful way.

6

i.   One-step residuals.

The GLS estimates of the regression parameters are obtained by performing the transformation

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

so that $E\left[\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon} \left(\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon} \right)' \right] = \sigma^2 \mathbf{I}$. The GLS transformation matrix for a regression model with AR(p) disturbances is well known and can be written as $\boldsymbol{\Omega} = h(\phi)$, where $h(\cdot)$ is a function given by the autoregression parameters (Gailbraith and Zinde-Walsh (10) supply exact and asymptotic expressions for a general stationary ARMA process). The transformation matrix for heteroscedastic variables can be of the class treated under Subsection ii below. The GLS estimates of $\mathbf{B}$ is then the usual OLS estimates of the transformed variables:

$$\hat{\mathbf{B}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}}'\tilde{\mathbf{Y}} \tag{4.3}$$

We will then define the estimate of $\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}$ as

$$\hat{\tilde{\boldsymbol{\varepsilon}}}_1 := \widehat{\boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}} = \tilde{\mathbf{Y}} - \hat{\tilde{\mathbf{Y}}} = \tilde{\mathbf{Y}} - \tilde{\mathbf{X}}\hat{\mathbf{B}} = \left(\mathbf{I} - \tilde{\mathbf{X}} \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}} \right)^{-1} \tilde{\mathbf{X}} \right) \tilde{\boldsymbol{\varepsilon}} \tag{4.4}$$

which can be considered to be one-step residuals as they result from one estimate. It is shown in Appendix A1 that, given normality, $\hat{\tilde{\boldsymbol{\varepsilon}}}_{1,A} \overset{!\bullet}{\rightarrow} \tilde{\boldsymbol{\varepsilon}}, \hat{\tilde{\boldsymbol{\varepsilon}}}_{1,H} \overset{!\bullet}{\rightarrow} \tilde{\boldsymbol{\varepsilon}}$ (where subscripts A and H denotes autocorrelation and heteroscedasticity). This fact ensures that the residuals are useful for non-normality testing.

ii.   Two-step residuals

Usually $\boldsymbol{\Omega}$ is unknown and have to be estimated. As this estimate generally is a function of the true unobservable disturbances, this has to be estimated first, hence the term *two-step* residuals.

Let us consider the situation with heteroscedastisity to begin with, where the structure of $\boldsymbol{\Omega}$ is a function of a subset of the regressors, i.e., $\boldsymbol{\Omega} = h(\mathbf{X}_2)$, where $\mathbf{X}_2 \subseteq \mathbf{X}$, e.g. $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$.   Let $\boldsymbol{\Omega} = \{\omega_{ii}\}_{i=1}^{n}$,   where

$\omega_{ii} = V\left(\varepsilon_{i,j}\right) = \mathbf{X}_{2,i}\theta$, $i = 1,2,...N$, $j = 1,2,...P$. Estimates of ω_{ii} can then be obtained from the linear regression model $\left(\boldsymbol{\varepsilon} \odot \boldsymbol{\varepsilon}\right) = \mathbf{X}_2\theta + \mathbf{v}$ (where $\mathbf{v}$ is an additive error term and $\odot$ is the Hadamard product) by the OLS estimates $\hat{\theta} = \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\left(\boldsymbol{\varepsilon} \odot \boldsymbol{\varepsilon}\right)$. Since $\theta_1 = ... = \theta_p$ by assumption, we can take $\hat{\Omega} = diag\left(\mathbf{X}_2\overline{\hat{\theta}}\right)$, where $\overline{\hat{\theta}} := \sum_{j=1}^{P}\hat{\theta}_j / P$. With $\boldsymbol{\varepsilon}$ replaced by $\hat{\boldsymbol{\varepsilon}}$ (the OLS residuals), our feasible estimate of Ω will be written as $\hat{\hat{\Omega}} = diag\left(\mathbf{X}_2\overline{\hat{\hat{\theta}}}\right)$. Using the estimated covariance matrix we can then obtain an estimate of $\tilde{\varepsilon}$ by

$$\hat{\hat{\varepsilon}}_{II,H} := \hat{\hat{\Omega}}^{-1/2}\hat{\varepsilon} = \hat{\hat{\Omega}}^{-1/2}\mathbf{M}_X\boldsymbol{\varepsilon} \tag{4.5}$$

It is shown in Appendix A2 that, given normality, $\hat{\hat{\varepsilon}}_{II,H} \xrightarrow{\ell_*} \tilde{\varepsilon}$.

The case of autocorrelation will be handled somewhat differently. Consider the AR(q) process

$$\varepsilon_{i,t} = \mathbf{Z}_{i,t}\phi_i + \delta_{i,t}, \quad i = 1,2,...,P. \tag{4.6}$$

where $\mathbf{Z}_{i,t} = \left[\varepsilon_{i,t-1} + ... + \varepsilon_{i,t-q}\right]$, $\phi_i = \left[\phi_{i,1},...,\phi_{i,q}\right]'$ and the roots of $\left(1 - \sum_{j=1}^{q}\phi_{i,j}z^j\right)$ lies outside the unit circle. The OLS estimate then becomes $\hat{\phi}_i = \left(\mathbf{Z}_{i,t}'\mathbf{Z}_{i,t}\right)^{-1}\mathbf{Z}_{i,t}'\varepsilon_{i,t}$. Again, since $\varepsilon_{i,t}$ is unobservable, we use $\hat{\varepsilon}_{i,t}$ to obtain the feasible OLS estimate $\hat{\hat{\phi}}_i = \left(\hat{\mathbf{Z}}_{i,t}'\hat{\mathbf{Z}}_{i,t}\right)^{-1}\hat{\mathbf{Z}}_{i,t}'\hat{\varepsilon}_{i,t}$. From this estimate we could achieve an estimated covariance matrix, and perform a transformation like that of (4.5). However, another type of residuals is frequently used in the literature, namely the so-called auxiliary regression residuals. It is therefore relevant to examine them. We define this variable as

$$\hat{\hat{\varepsilon}}_{II,i,A} := \hat{\delta}_{i,A} = \hat{\varepsilon}_{i,t} - \hat{\hat{\varepsilon}}_{i,t} = \hat{\varepsilon}_{i,t} - \hat{\varepsilon}_{i,t-1}\hat{\hat{\phi}}_1 + ... + \hat{\varepsilon}_{i,t-q}\hat{\hat{\phi}}_q. \tag{4.7}$$

It may be shown that, given normality, $\hat{\hat{\varepsilon}}_{II,i,A} \xrightarrow{\ell_*} \delta_i$ (Appendix A2). The argumentation above is easily extended to an arbitrary q. Consequently, $\hat{\hat{\varepsilon}}_{II,A} := \left[\hat{\hat{\delta}}_{II,1,A} \ \hat{\hat{\delta}}_{II,2,A} \cdots \hat{\hat{\delta}}_{II,P,A}\right]$ can be used to examine the distribution of δ_A. As (4.5) and (4.7) result from estimating $\boldsymbol{\varepsilon}$ first from the OLS estimate and then estimating ϕ and θ from the OLS residuals, we will refer to these as *two step residuals*, thus the index II.

iii.   Three-step residuals

The two-step residuals mentioned above are perhaps not the most intuitive to use, at least not for the case of heteroscedasticity. Rather, one may employ the residuals from the transformed FGLS model. Using the estimated covariance matrixes treated in subsection (ii), we may perform the transformation

$$\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon} \rightarrow \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{Y} = \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{X} + \hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{\varepsilon} , \text{ or, } \widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\mathbf{B} + \widetilde{\boldsymbol{\varepsilon}} \tag{4.8}$$

The estimated covariance matrix for the case of heteroscedasticity is given in (ii). For multivariate FGLS to operate in the case of autocorrelation we assume $\phi_1 = \phi_2 = ... = \phi_P$. We then take the mean value of the estimates above as our final estimate: $\overline{\hat{\phi}} = \sum_{i=1}^{P} \hat{\phi}_i \big/ P$. The residuals of (4.8) is then

$$\hat{\tilde{\boldsymbol{\varepsilon}}}_{III} := \widetilde{\mathbf{Y}} - \widetilde{\widetilde{\mathbf{Y}}} = \left(\mathbf{I} - \hat{\boldsymbol{\Omega}}^{-1/2}\mathbf{X} \left(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X} \right) \mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1/2} \right) \hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{\varepsilon} = \mathbf{M}_{\hat{\boldsymbol{\Omega}}^{-1/2}}\hat{\boldsymbol{\Omega}}^{-1/2}\boldsymbol{\varepsilon} \tag{4.9}$$

Given normality it may be shown that $\hat{\tilde{\boldsymbol{\varepsilon}}}_{III,H} \xrightarrow{l\cdot} \tilde{\boldsymbol{\varepsilon}}$, $\hat{\tilde{\boldsymbol{\varepsilon}}}_{III,A} \xrightarrow{l\cdot} \tilde{\boldsymbol{\varepsilon}}$ (Appendix A3). These vectors results from first calculating the OLS residuals, then estimating the covariance matrix, and finally calculate the FGLS residuals. We will therefore refer to these as *three step residuals*.

The rate of convergence of the one, two- and three step residuals mentioned above are likely to differ. Also, they are all functions of the regressors, so the distributional properties of the regressors will affect the distributional properties of the proxy variables. In addition one may expect that they are unequally robust to misspecifications of the autocorrelation or heteroscedasticity. We will therefore explore and compare the properties of the residuals of Section **IV** by means of Monte Carlo simulations, where factors such as distribution of regressors and covariance structures are varied. We will discuss this further on in Section **VI**.

V. The Jarque&McKenzie non-normality test.

Tests of marginal normality do not examine multivariate normality as they ignore the correlations between the variables. Consequently, genuine tests for multivariate normality are needed in order to assess multivariate normality. As we are merely interested in the properties of the residuals´ usefulness to test for non-normality, rather than the properties of various tests, we will choose one single non-normality test with known (good) properties. One such test is the Jarque and Mckenzie (3) test statistic (JM), which is the sum of Mardia (11) measures of kurtosis and skewness defined as

$$\gamma_{1,P} = E\left[\left(X-\mu\right)'\Sigma^{-1}\left(X-\mu\right)\right]^3, \qquad \gamma_{2,P} = E\left[\left(X-\mu\right)'\Sigma^{-1}\left(X-\mu\right)\right]^2.$$

Estimates of these measures are

$$\hat{\gamma}_{1,P} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\left(X_i-\bar{X}\right)'S^{-1}\left(X_j-\bar{X}\right)\right]^3, \qquad \hat{\gamma}_{2,P} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(X_i-\bar{X}\right)'S^{-1}\left(X_i-\bar{X}\right)\right]^2,$$

with

$$D_1 = \frac{N\hat{\gamma}_{1p}}{6} \xrightarrow{\ell} \chi^2_{(P(p+1)(p+2)/6)}, \qquad D_2 = \frac{\sqrt{N}\left(\hat{\gamma}_{2p}-P(P+2)\right)}{\sqrt{8P(P+2)}} \xrightarrow{\ell} N(0,1),$$

where N is the number of observations. Jarque and Mckenzie (3) suggested the omnibus test

$$D_P(X) = D_1(X) + D_2^2(X), \text{ where } D_P(X) \xrightarrow{\ell} \chi^2_{1+p(p+1)(p+2)/6}. \qquad (5.1)$$

The null hypothesis $H_0: X \sim N(0,\Sigma,\Xi)$ is then rejected at the α-level whenever

$$D_P(X) > \eta \text{ where } P\left(D_P(X) > \eta \mid \gamma_{1,P} = 0 \cap \gamma_{2,P} = P(2+P)\right) = \alpha.$$

Note that, according to (3.1) and (3.6), we need $\Xi = \sigma^2 I$ for (5.1) to hold. As the limiting null distribution of any of the residuals $\hat{\varepsilon}$ in Section **IV** equals that of the true

10

disturbances, $D_P\left(\hat{\tilde{\varepsilon}}\right)$ can be used in order to test for non-normality. However, the small sample properties of $\hat{\tilde{\varepsilon}}$ will depend on **X**. It is shown in Holgersson and Shukur (1) that the size of the JM test might be biased in small samples, due to the properties of **X**. In addition, the JM statistic itself converges rather slowly to its asymptotic null distribution. But when Ω is known, we can actually do better than base the JM test directly on the GLS residuals. As the JM statistic is invariant to linear transformations, the size may be controlled by Monte Carlo techniques in the following way: Let T be a statistic such that the null hypothesis, H_0: $\tilde{\varepsilon} \sim N\left(0,\Sigma,\sigma^2\mathbf{I}\right)$, is rejected when $T \geq c$, and let $G(x)=P[T \geq x]$ so that $G(c)=\alpha$. Then define

$$\tilde{G}_r(x)=\#\{T_i \geq x\}/r, \quad \hat{G}_r(x)=\frac{r\tilde{G}_r(x)+1}{r+1}, \text{ where } T_i \ (i=1, 2,...,\text{r}) \text{ is an independent}$$

realisation of T (i.e., a Monte Carlo replicate under the null hypothesis). The associated (Monte Carlo) critical region is then defined as $\hat{G}_r\left(T_0\right) \leq \alpha$, so that $\hat{G}_r\left(T_0\right)$ may be interpreted as an estimate of $G\left(T_0\right)$. If $\alpha(r+1)$ is an integer, we have $P\left[\hat{G}_r\left(T_0\right) \leq \alpha\right]=\alpha$ (see e.g. Dufour, *et. al.* (12)). As ε is unobservable, we can use the observable linear combination $\hat{\tilde{\varepsilon}}_1:=\left(\mathbf{I}-\tilde{\mathbf{X}}\left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)\tilde{\mathbf{X}}\right)\tilde{\varepsilon}=\mathbf{M}_{\tilde{\mathbf{X}}}\tilde{\varepsilon}$, i.e. our one-step residuals, to obtain a Monte Carlo critical region for $G\left(D_P\left(\varepsilon\right)\right)$ by simply taking $T_0=\mathbf{M}_{\tilde{\mathbf{X}}}\tilde{\varepsilon}$ and $T_i=\mathbf{M}_{\tilde{\mathbf{X}}}\delta_i$, where $\delta_i \sim N_{P,N}\left(0,\ \mathbf{I}_{(P\times P)},\ \mathbf{I}_{(N\times N)}\right)$, since δ has the same distribution as $\tilde{\varepsilon}$ under the null hypothesis. The null hypothesis H_0: $\varepsilon \sim N\left(0,\Sigma,\Xi\right)$ is then rejected at the α-level whenever $\hat{G}_r\left(D_{P,0}\left(\tilde{\varepsilon}\right)\right)<\alpha$, i.e.

$$P\left(\hat{G}_N\left(D_{P,0}\left(\tilde{\varepsilon}\right)\right)>\alpha\mid\gamma_{1,P}=0\cap\gamma_{2,P}=P(2+P)\right)=\alpha.$$

By using this approach for the GLS residuals we obtain an exact inference procedure that will be equivalent to using the true unobservable disturbances in conjunction with exact critical values for the statistic.

11

VI. The Monte Carlo design

In this section we will 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 **IV**. First, we wish to see if the actual size of the test (i.e. the probability of rejecting the null when true) is close to the nominal size. Since the experiment is performed using a finite number of replicates, we need to be able to distinguish simulation fluctuations from biases in the test. One possibility to do this is to calculate an approximate 95% confidence interval for the actual size α:

$$\hat{\alpha} \pm 1.96 \sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{R}}.$$

where $\hat{\alpha}$ is the estimated size and R is the number of replicates. To judge the reasonability of the results, we require that the estimated size should lie within the 95% confidence interval. For example, if we consider a nominal size of 5%, and we operate 10 000 replicates, we define a result as reasonable if the estimated size lies between 0.0457 and 0.0543.

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 skewed and one symmetric alternative distribution. The first is defined by $\mathbf{X} = \mathbf{LY}$ where Y_i (i = 1, 2, ...,P) are iid $\chi^2_{(v)}$ variates, and $\mathbf{LL'} = \Sigma_{P \times P}$, a covariance matrix. We shall write this distribution as $\chi(v, \Sigma)$. As we are examining a test that only uses information of the skewness and kurtosis, it is of relevance to consider the power to detect distributions with marginal normal skewness and kurtosis. We therefore make use of the Khintchine distribution to generate our symmetric distribution. The variable is defined by $X_i = \pi_i R_i U$, $i = 1, 2, ...P$, where $U_i \sim U[0,1]$, $R_i \sim [\Gamma(\lambda, \varsigma)]^\tau$ and

$$p(\pi_i) = \begin{cases} 0.5 & \pi_i = 1 \\ 0.5 & \pi_i = -1 \end{cases}.$$

We will denote the Khintchine variable as $K_{(\tau)}$. In order to invoke covariance we will use $R_1 = R_2 = \ldots = R_P$ (and $U_1 \neq U_2 \neq \ldots \neq U_P$). The coefficient of kurtosis of this distribution is determined by the expression $\beta_2 = \dfrac{\Gamma(\varsigma + 4\tau)\Gamma(\varsigma)}{\Gamma(\varsigma + 2\tau)^2}$ (Johnson (13)). In this study we choose $\tau = 0.1$ with the corresponding value $\varsigma = 0.12757$ that yields $\beta_2 = 3$ (the marginal kurtosis of a normal distribution).

In order to examine the effects of autocorrelated noise, we need to choose some forms of autocovariances. For simplicity, we consider (marginal) AR processes of low orders. As it is not likely that the true data generating process is known, we will examine the consequence of misspecifying the autocovariance structure of the disturbances. The effects of heteroscedastisity will be examined using covariance matrices that are a function of a subset of the regressors, as on page 7. Again, we will examine the case where the covariance structure is correct as well as the case when it is misspecified.

An additional aspect of relevance is the consequence of using the observable linear combination $\mathbf{M}\boldsymbol{\varepsilon}$ as a proxy to $\boldsymbol{\varepsilon}$. In Holgersson and Shukur (1) the simulation results indicate that the biases in the estimated skewness and kurtosis, due to $\mathbf{M}$, 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. This effect is expected to be worse whenever the disturbances are non-spherical, as the FGLS residuals are a much more complicated function of $\mathbf{X}$ than the OLS residuals is. Therefore, the interaction effect of the distribution of $\mathbf{X}$ and the structure of the autocorrelation is of great relevance and will hence be examined in the experiment. In order to obtain such heavy tailed regressors we use a variable defined by $\mathbf{X} = \mathbf{LT}$, where T_i (i = 1, 2,…,P) are iid $t_{(v)}$ distributed variates, and $\mathbf{LL'} = \Sigma_{P \times P}$. We will write this distribution as T(v, Σ).

Below we present a table of all the factors treated in the experiment.

TABLE 1. Properties of factors in the experiment.

Factor	Symbol	
Test variables	$\hat{\varepsilon}_{\mathrm{I}}$	One-step residuals
	$\hat{\varepsilon}_{\mathrm{II}}$	Two-step residuals
	$\hat{\varepsilon}_{\mathrm{III}}$	Three-step residuals
Numbers of regressors	k	7
Distribution of regressors	$\mathbf{X}_{[1:5]}$	$T(0, 1, \Sigma_{\mathbf{X}})$, $N(0, \Sigma_{\mathbf{X}})$ *
	$\mathbf{X}_{[6:7]}$	$U(1, 10)$
Nominal size	α	5%
Number of equations	P	5
Number of degrees of freedom (P(N-k-1))	ν	15, 35, 75, 100, 150, 200, 400.
Distribution of disturbances	$\varepsilon \sim$	$N(0, \Sigma_{\varepsilon})$, $K_{(0.1)}$, $\chi(3, \Sigma_{\varepsilon})$. **
Order of AR process	q	1, 3
Structure of Heteroscedstisity	Ω	$diag\left(\mathbf{X}_{[6:7]}\theta_{(2\times 1)}\right)$
Value of parameter in the AR(1) processes	ϕ	0.8
Value of parameters in the AR(3) processes	$[\phi_1\ \phi_2\ \phi_3]$	0.8, 0.08, 0.04
Value of Heteroscedasticity parameters	$[\theta_1\ \theta_2]'$	$[2\ 5]'$

* $\Sigma_{\mathbf{X}} = toeplitz(5\ 3\ 1.8\ 1.2\ 0.2)$.

** $\Sigma_{\varepsilon} = toepliz(25\ 15\ 7\ 1\ 0.5)$.

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 JM test. The results will be presented in graphical forms in two parts, *size properties* and *power properties* respectively. All simulations have been performed using R = 10 000 replicates. Each graph consists of results of three different test procedures: the one-step residuals of **IV:i** have been used along with Monte Carlo critical region (denoted by), the two-step residuals of **IV:ii** have been used with the critical region defined by the asymptotic chi-2 distribution (denoted by Υ), and finally, the three-step residuals of **IV:iii** as test variable, again using the asymptotic chi-2 distribution (denoted by). The   lines will thus represent the exact critical region conditioned on **X**, while the lines marked by Υ and   will represent critical regions unconditioned on **X**, although applied to two different test variables. The Monte Carlo critical region is decided from r = 59 Monte Carlo replications. Finally, the autoregressive processes have been generated using 30 "start-up" values.

Size properties

In this subsection we present our results concerning the size properties of the JM test when applied to the various residuals of section **IV**. Figure 1 shows the size properties of the JM test when the regressors are following a heavy tailed distribution. There is a clear distinction between the variables in the sense that the two-step residuals converge much slower than the three-step residuals. Figure 2 suggests that a small misspecification of the autoregressive order does not alter the size properties of the test materially for the two- and three-step residuals. In fact, the size seems to limit the nominal size, regardless of the misspecification. Figure 3 visualises the impact of a small misspecification of the autoregressive order in conjunction with heavy-tailed regressors, which is a fairly realistic situation. On comparison with Figure 2, we see that the distribution of the regressors causes the test to diverge for the auxiliary regression residuals, while the behaviour of the FGLS residuals remains unchanged from Figure 2. This indicates that even though the two types of residuals are asymptotically equivalent under correct specifications, they may have totally different

15

properties when there is a small misspecification. In figure 4 we see that the two- and three step residuals causes the test to overreject, though the test based on the two-step residuals are more close to its nominal size, a result in line with that of Figure 2 and Figure 3. When we impose a misspecification of the heteroscedasticity, as in Figure 5, the size clearly diverges. As for the autocorrelation case, heavy tailed regressors seem to worsen the performance according to Figure 6. In general, the three-step residuals perform better than the two-step residuals, especially for misspecified autocorrelation in conjunction with heavy-tailed regressors, in which case the test diverges.

From Figure 1-6, we see that it is indeed possible to obtain exact size by using the $\mathbf{M}$ matrix when determining the critical region, whenever $\mathbf{\Omega}$ is known. 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 covariance matrixes do exist in some situations (indeed, this is the situation examined in most empirical studies of non-normality tests).

Figure 1. *The estimated size for the JM test at 5% level, disturbances defined by* $\varepsilon_t = \delta_t$, *though specified as* $\varepsilon_t = \phi\varepsilon_{t-1} + \delta_t$ *(i.e., over specified autocorrelation) and* $X \sim T(1, \Sigma_X)$.
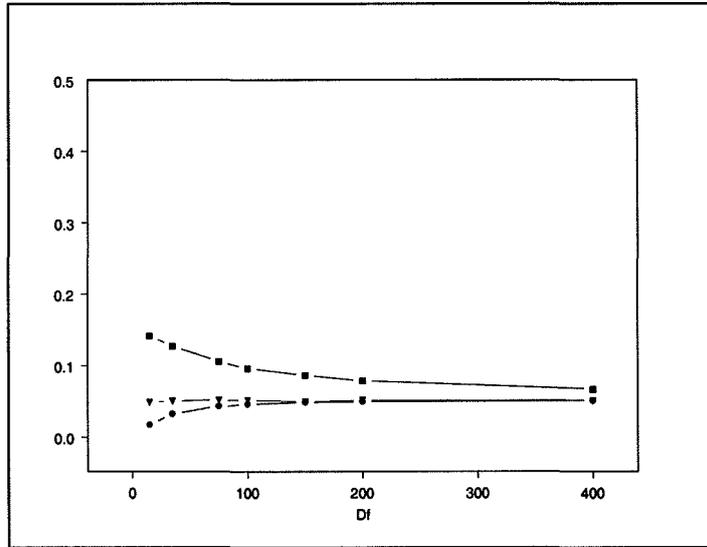
Figure 2. *The estimated size for the JM test at 5% level, disturbances defined by* $\varepsilon_t = 0.8\varepsilon_{t-1} + 0.08\varepsilon_{t-2} + 0.04\varepsilon_{t-3} + \delta_t$, *though specified as* $\varepsilon_t = \phi\varepsilon_{t-1} + \delta_t$ *(i.e., under specified autocorrelation) and* $X \sim N(0, \Sigma_X)$.
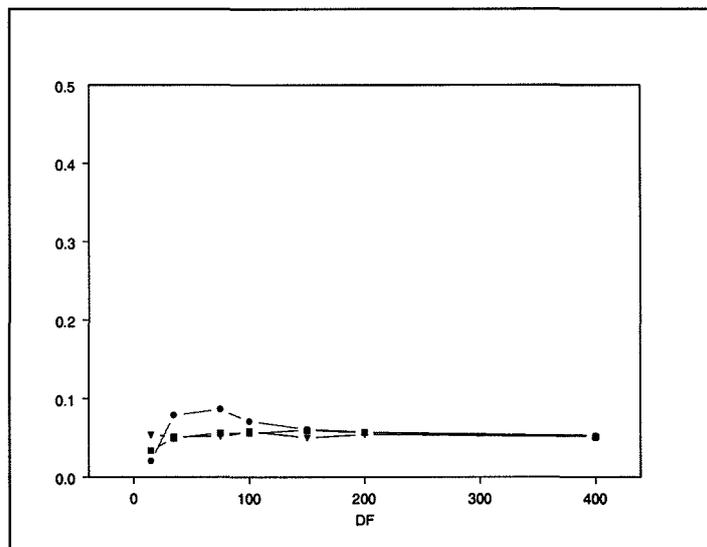
Figure 3. *The estimated size for the JM test at 5% level, disturbances defined by* $\varepsilon_t = 0.8\varepsilon_{t-1} + 0.08\varepsilon_{t-2} + 0.04\varepsilon_{t-3} + \delta_t$, *though specified as* $\varepsilon_t = \phi\varepsilon_{t-1} + \delta_t$ *(i.e., under specified autocorrelation) and* $X \sim T(0, 1, \Sigma_X)$.
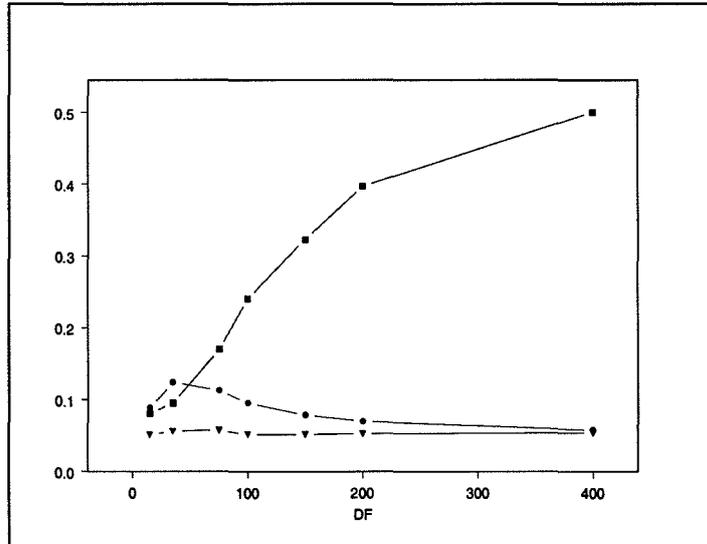
Figure 4. *The estimated size for the JM test at the 5% level based on the FGLS residuals, where* $\Omega = diag\left(\left[X_{6,7}\right]\left[2 \quad 5\right]'\right)$, *specified as* $\Omega = diag\left(\left[X_{6,7}\right]\left[\theta_1 \quad \theta_2\right]'\right)$ *(i.e., correct specified heteroscedasticity) and* $X \sim N(0, 1, \Sigma_X)$ *distributed regressors.*
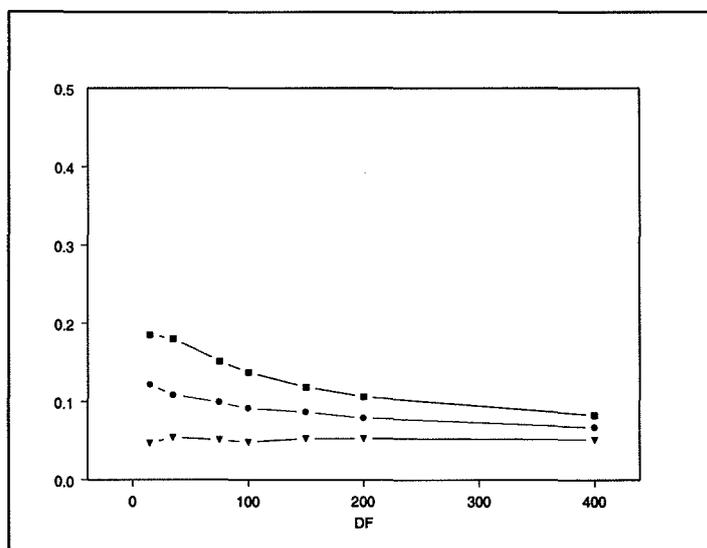
18

Figure 5. *The estimated size for the JM test at the 5% level based on the FGLS residuals, where* $\Omega = diag\left(\begin{bmatrix} \mathbf{X}_6 & \mathbf{X}_7 \end{bmatrix} \begin{bmatrix} 2 & 5 \end{bmatrix}' \right)$, *specified as* $\Omega = diag\left(\mathbf{X}_6 \theta_1 \right)$ *(i.e., miss-specified heteroscedasticity) and* $\mathbf{X} \sim N\left(\mathbf{0}, \Sigma_{\mathbf{X}} \right)$ *distributed regressors.*
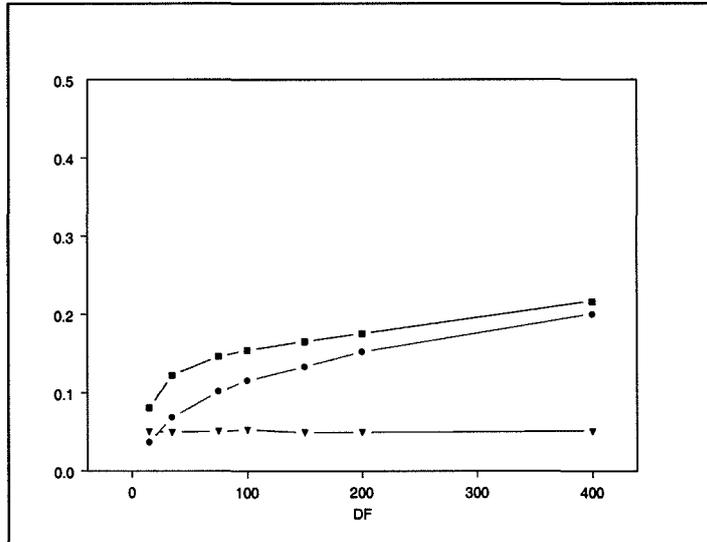
Figure 6. *The estimated size for the JM test at the 5% level based on the FGLS residuals, where* $\Omega = diag\left(\begin{bmatrix} \mathbf{X}_6 & \mathbf{X}_7 \end{bmatrix} \begin{bmatrix} 2 & 5 \end{bmatrix}' \right)$, *specified as* $\Omega = diag\left(\mathbf{X}_6 \theta_1 \right)$ *(i.e., miss-specified heteroscedasticity) and* $\mathbf{X} \sim T\left(\mathbf{0}, 1, \Sigma_{\mathbf{X}} \right)$ *distributed regressors.*

19
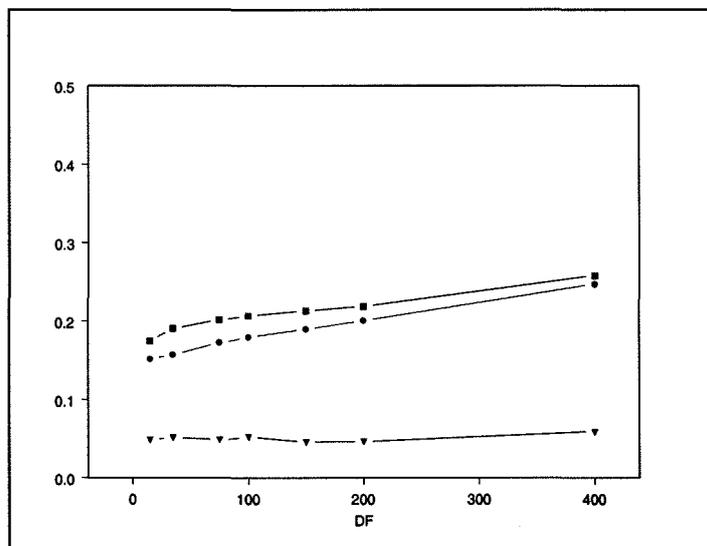
Power properties

In this subsection we present our results concerning the power properties of the JM test under various situations that are likely to appear in applied studies. Figure 7 reveals that the power of the test is fairly high for all three types of residuals when the target variable ε is following the skew distribution. As expected, the one-step procedure performs better in a comparison between Figure 7 and Figure 8; the effect of underdimensioned order of the autoregressive process causes the power to be lower. Also, the test based on the one-step residuals have slightly better power as compared to that of the three-step residuals, which in turn is more powerful than that of the two-step residuals. Figure 9 shows the power of the test for the Khintchine variable, which is fairly high for all three residuals, higher for the one-step residuals though. A comparison between Figures 9 and 10 shows that an under specification of the heteroscedasticity does not appear to lower the power markedly. Moreover, the heavy tailed regressors do not seem to cause a serious reduction of the power. Thus, estimating the covariance matrix matters in small samples in the sense that it will lead to a power reduction.

Figure 7. *The estimated power for the JM test at the 5% level, disturbances defined by* $\varepsilon_t = \delta_t$, *though specified as* $\varepsilon_t = \phi\varepsilon_{t-1} + \delta_t$ *(i.e., over specified autocorrelation) with* $\mathbf{X} \sim N(\mathbf{0}, \Sigma_{\mathbf{X}})$, $\delta_t \sim \chi^2(3, \Sigma_\delta)$.
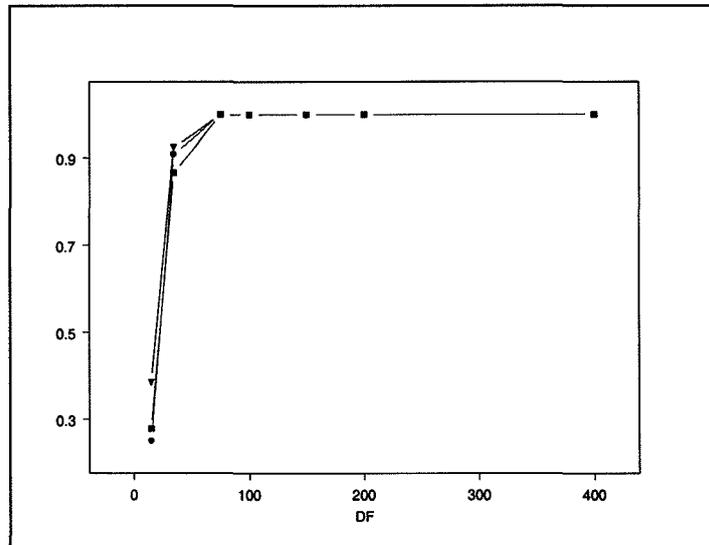
Figure 8. *The estimated power for the JM test at the 5% level, disturbances defined by* $\varepsilon_t = 0.8\varepsilon_{t-1} + 0.08\varepsilon_{t-2} + 0.02\varepsilon_{t-3} + \delta_t$ *though specified as* $\varepsilon_t = \phi\varepsilon_{t-1} + \delta_t$ *(i.e., under specified autocorrelation) with* $\mathbf{X} \sim N(\mathbf{0}, \Sigma_{\mathbf{X}})$, $\delta_t \sim \chi^2(3, \Sigma_\delta)$.

21
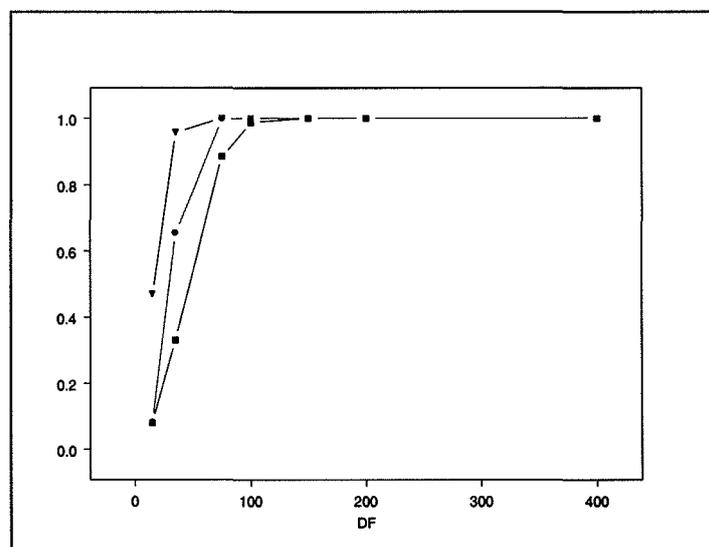
Figure 9. *The estimated power for the JM test at the 5% level based on the FGLS residuals, where* $\Omega = diag\left(\left[\mathbf{X}_{6,7}\right]\left[2 \quad 5\right]'\right)$*, specified as* $\Omega = diag\left(\left[\mathbf{X}_{6,7}\right]\left[\theta_1 \quad \theta_2\right]'\right)$ *(i.e., correct specified heteroscedasticity) and* $\mathbf{X} \sim T\left(0, 1, \Sigma_{\mathbf{X}}\right)$ *distributed regressors and* $\delta \sim K_{(0.1)}$.
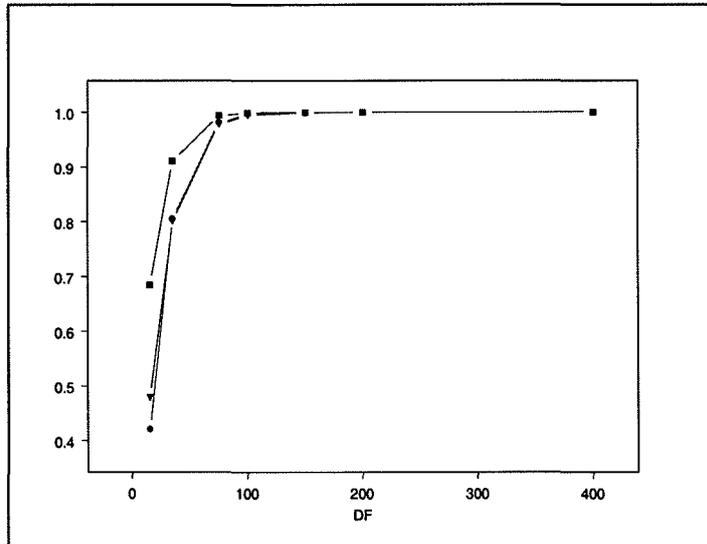
Figure 10. *The estimated power for the JM test at the 5% level based on the FGLS residuals, where* $\Omega = diag\left(\left[\mathbf{X}_6 \quad \mathbf{X}_7\right]\left[2 \quad 5\right]'\right)$*, specified as* $\Omega = diag\left(\mathbf{X}_6\theta_1\right)$ *(i.e., miss-specified heteroscedasticity) and* $\mathbf{X} \sim T\left(0, 1, \Sigma_{\mathbf{X}}\right)$ *distributed regressors and* $\delta \sim K_{(0.1)}$.

22
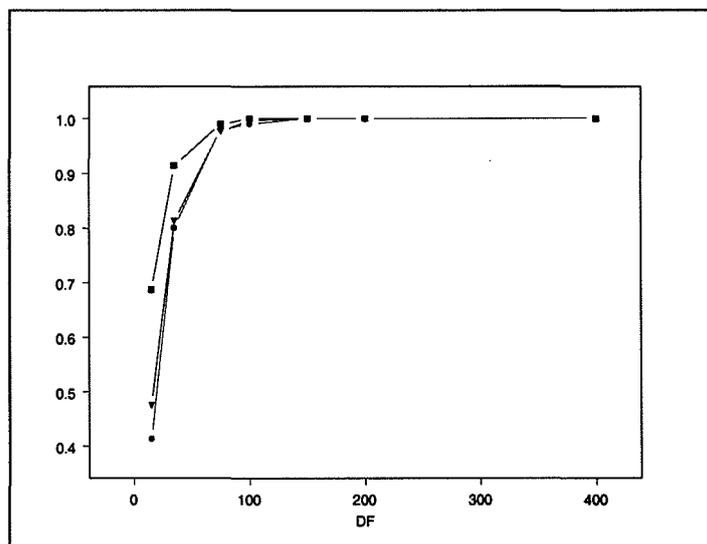
VIII. Conclusions and summary

In this paper we have studied the effect of using the JM test for non-normality in the presence of autocorrelation or heteroscedasticity (i.e. *identically dependent distributed* disturbances and *independent heterogeneously distributed* disturbances) when applied to multivariate regression models. We have motivated why moment based tests in general will not be valid for heteroscedastic or autocorrelated variables. A number of models were investigated in order to enlighten the effect of some frequently occurring properties of real data, such as heavy-tailed regressors along with known/unknown autoregressive order or known/unknown heteroscedasticity. In addition we have shown that for the case of known covariance matrix (GLS), a simple Monte Carlo method can be used in order to obtain a test with exact size, regardless of the number of observations, equations or properties of the regressors. For each model we have performed 10 000 replications, varying sample sizes ranging from 15 to 400 degrees of freedom. In addition, the power properties have been examined for one skewed and one symmetric distribution. The simulations revealed that even a small misspecification of the autoregressive order or the heteroscedasticity may ruin the test if the regressors are heavy-tailed, in the sense that the size will be far above the nominal size, as the effect increases with increasing sample size. Thus the use of regular non-normality tests on variables with a complicated data generating process, such as in economic applications, is dubious. However, it may still be informative to apply the JM test (or other non-normality tests) in non-experimental situations. Indeed, if we do not reject the null hypothesis of multivariate normality, it is plausible that the disturbances are not just normally distributed, but iid normal, which in turn is a strong indication that our modeling is successful. On the contrary, if we reject the null hypothesis, this may very well be due to other causes than non-normality. In addition, the power properties seem to be reduced when there is a small misspecification of the autoregressive order. This indicates that even though the two types of residuals are asymptotically equivalent under correct specifications, they may have totally different properties when there is a misspecification.

23

Appendix: Asymptotic distribution of I, II and III-step residuals.

Before we examine the asymptotic properties of our residuals of Section **IV**, we will list some properties of the so-called hat matrix that are crucial for our analysis. Consider a regression model $Y = Z\beta + \varepsilon$. The hat matrix of a regression estimate is defined as $H := Z(Z'Z)^{-1}Z'$. The OLS residuals can then be expressed as $\hat{\varepsilon} := Y - \hat{Y} = (I - H)\varepsilon$. As H is symmetric and idempotent it follows that $\hbar(H) = k$, $0 \le h_{ii} \le 1$. Assuming that $V(\varepsilon) = \sigma^2 I$, i.e. that the variance is scalar, the variance of the difference between the residuals and the disturbances can be expressed as $V(\xi) := V(\hat{\varepsilon} - \varepsilon) = \sigma^2 H$, or, $V(\xi_i) = V(\hat{\varepsilon}_i - \varepsilon_i) = \sigma^2 h_{ii}$. By Chebychev's inequality

$$P(|\xi_i - E(\xi_i)| \ge v) = P(|\hat{\varepsilon}_i - \varepsilon_i| \ge v) \le \frac{h_{ii}\sigma^2}{v^2}. \text{ Thus} \max_{1 \le i \le n}(h_{ii}) \to 0 \Rightarrow |\hat{\varepsilon}_i - \varepsilon_i| \overset{p}{\to} 0.$$ The

event $\max_{1 \le i \le n}(h_{ii}) \to 0$ is known as the Huber (14) condition. Since the i:th diagonal element of H can be written as $h_{ii} = (X_{i1},...,X_{ik})(X'X)^{-1}(X_{i1},...,X_{ik})'$, and k is a fixed finite number, it follows that $plim(X'X)^{-1} = 0$ suffices for the Huber condition to hold. It can be shown that our assumptions in (1.1) implies $plim(X'X)^{-1} = 0$ (Judge, *et. al.* (15)).

We will apply these properties of the hat-matrix in order to analyse the asymptotic properties of our residuals of Section **IV**.

A1. Limit of one-step residuals.

The one-step residuals are $\hat{\tilde{\varepsilon}}_I := \left(\tilde{Y} - \hat{\tilde{Y}}\right) = \left(I - H_{(\tilde{x})}\right)\tilde{\varepsilon}$ for $H_{\tilde{x}} := \tilde{X}\left(\tilde{X}'\tilde{X}\right)\tilde{X}'$. The limiting properties of these will be analysed separately for the autocorrelated and the heteroscedastic disturbances:

Heteroscedasticity:

The variance of the difference of the one-step residuals and the transformed disturbances are $V\left(\hat{\tilde{\varepsilon}}_{I,H} - \tilde{\varepsilon}\right) = \sigma^2 H_{(\tilde{x})}$. Thus $\max_{1 \leq i \leq n}\left(h_{(\tilde{x})ii}\right) \rightarrow 0$ suffices for $\hat{\tilde{\varepsilon}}_i \overset{p}{\rightarrow} \tilde{\varepsilon}_i$ to hold, which we have from our assumptions of (1.1) p.2 and from p.24. Finally we have from Rao (16) $\left|X_n - Y_n\right| \overset{p}{\rightarrow} 0$, $Y_n \overset{p}{\rightarrow} Y \Rightarrow X_n \overset{\ell}{\rightarrow} Y$, hence $\hat{\tilde{\varepsilon}}_{i,I,H} \overset{\ell}{\rightarrow} \tilde{\varepsilon}_i$, so that $\hat{\tilde{\varepsilon}}_{I,H} \overset{\ell}{\rightarrow} \tilde{\varepsilon}$.

Autocorrelation:

We study the special case with AR(1) process. Recall that by assumption $|\phi| < 1$. Let

$$\hat{\tilde{\varepsilon}}_{I,A} := M_{(\tilde{x})}\Omega^{-1/2}\varepsilon \text{ where } M_{(\tilde{x})} \equiv I - \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}' = I - H_{(\tilde{x})}.$$

Unlike the case of heteroscedasticity, it does not make sense to assume that $\varepsilon = \Omega^{1/2}\delta$ where δ is some iid variable. Rather, we consider ε as generated from the model $\varepsilon_t = \phi\varepsilon_{t-1} + \delta_t$ so that $E[\varepsilon\varepsilon'] = \Omega$, which means that convergence in probability cannot be shown as above. To determine the asymptotical distribution of $\hat{\tilde{\varepsilon}}_{I,A}$ we examine the limiting moments of $\hat{\tilde{\varepsilon}}_{I,A} - \delta$ under the null hypothesis of normality:

i. $E\left[\hat{\tilde{\varepsilon}}_i - \delta_i\right] = 0 - 0 = 0$

ii. $V\left[\hat{\tilde{\varepsilon}}_i - \delta_i\right] = \sigma^2 h_{(\tilde{x})ii}$, i.e. $\max\left(h_{(\tilde{x})ii}\right) \overset{p}{\rightarrow} 0 \Rightarrow V\left[\hat{\tilde{\varepsilon}}\right] \overset{p}{\rightarrow} V[\delta] = \sigma^2$. As $\left\{\delta_{i,j-s}\right\} s \in \mathbb{Z}$ is normally distributed, ε_{ij} will be normal as well, though with other moments. Because the normal distribution is completely specified from its first two central moments it follows from above that $\hat{\tilde{\varepsilon}}_{ij} \overset{\ell}{\rightarrow} \delta_{ij}$, or, $\hat{\tilde{\varepsilon}} \overset{\ell}{\rightarrow} \delta$.

Appendix A2: Two-step residuals.

Heteroscedasity:

A sufficient condition for the residuals to converge in law to the disturbances is that their difference converges in probability to zero. This difference can be written as $\hat{\tilde{\varepsilon}}_{II,H} - \tilde{\varepsilon} = \hat{\Omega}^{-1/2}\mathbf{M}_{(X)}\varepsilon - \Omega^{-1/2}\varepsilon$, or,

$$
\begin{pmatrix} \hat{\omega}^{11} & 0 & \cdots \\ \vdots & \hat{\omega}^{22} & \\ & & \ddots \end{pmatrix}\begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \end{pmatrix} - \begin{pmatrix} \omega^{11} & 0 & \cdots \\ \vdots & \omega^{22} & \\ & & \ddots \end{pmatrix}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} \hat{\omega}^{11}\hat{\varepsilon}_1 \\ \hat{\omega}^{22}\hat{\varepsilon}_2 \\ \vdots \end{pmatrix} - \begin{pmatrix} \omega^{11}\varepsilon_1 \\ \omega^{22}\varepsilon_2 \\ \vdots \end{pmatrix}
$$

where ω^{ii} is the i:th diagonal element of $\Omega^{-1/2}$. Since $\hat{\theta}_i$ is consistent (e.g. Amemiya (17)), it follows that $\overline{\hat{\theta}}$ is consistent as well, and from the Slutsky theorem we get $X_{2,i}\overline{\hat{\theta}} \xrightarrow{p} X_{2,i}\theta = \omega_i$, i.e. $\hat{\omega}^{ii} \xrightarrow{p} \omega^{ii}$, and from p. 24 we have $\hat{\varepsilon}_i \xrightarrow{p} \varepsilon_i$. Thus $\hat{\omega}^{ii}\hat{\varepsilon}_i - \omega^{ii}\varepsilon_i \xrightarrow{p} \omega^{ii}\varepsilon_i - \omega^{ii}\varepsilon_i = 0$, i.e. $\hat{\tilde{\varepsilon}}_i \xrightarrow{\ell} \tilde{\varepsilon}_i$, and so $\hat{\tilde{\varepsilon}} \xrightarrow{\ell} \tilde{\varepsilon}$.

Autocorrelation:

That the limiting distribution of these residuals equals that of δ in (4.6) can be seen by expanding $\hat{\delta}_{i,A}$ for the special case $q = 1$ (for an AR(1) process):

$$
\delta_t - \hat{\delta}_t = \delta_t - \left(\hat{\varepsilon}_{i,t} - \hat{\hat{\varepsilon}}_{i,t}\right) = \delta_t - \mathbf{M}_{X_t}\varepsilon_{i,t} + \mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\hat{\phi} =
$$

$$
\delta_t - \mathbf{M}_{X_t}\left[\varepsilon_{i,t-1}\phi + \delta_{i,t}\right] + \mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\left[\left(\hat{\varepsilon}'_{i,t-1}\hat{\varepsilon}_{i,t-1}\right)^{-1}\hat{\varepsilon}'_{i,t-1}\hat{\varepsilon}_{i,t}\right] =
$$

$$
\delta_t - \mathbf{M}_{X_t}\varepsilon_{i,t-1}\phi - \mathbf{M}_{X_t}\delta_{i,t} + \mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\left[\left(\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\right)^{-1}\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\mathbf{M}_{X_t}\left\{\varepsilon_{i,t-1}\phi + \delta_{i,t}\right\}\right] =
$$

$$
\delta_t - \mathbf{M}_{X_t}\delta_{i,t} - \mathbf{M}_{X_t}\varepsilon_{i,t-1}\phi + \mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\left(\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\right)^{-1}\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\mathbf{M}_{X_t}\varepsilon_{i,t-1}\phi +
$$

$$
\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\left(\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\right)^{-1}\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\mathbf{M}_{X_t}\delta_{i,t} =
$$

$$
\left(I - \mathbf{M}_{X_t}\right)\delta_{i,t} - \left(I - \mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\left(\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\right)^{-1}\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\right)\mathbf{M}_{X_t}\varepsilon_{i,t-1}\phi +
$$

$$
\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\left(\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\varepsilon_{i,t-1}\right)^{-1}\varepsilon'_{i,t-1}\mathbf{M}_{X_{t-1}}\mathbf{M}_{X_t}\delta_{i,t} \tag{A2.1}
$$

Lemma:

Let $\mathbf{X}_{(n \times k)}$ and $\mathbf{Z}_{(n \times P)}$ be random matrices such that $E[\mathbf{X}'\mathbf{Z}] = 0$ and $\mathbf{M}_{(\mathbf{x})} := \left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) =: \left(\mathbf{I} - \mathbf{H}_{(\mathbf{x})}\right)$ where $p\lim(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{Q}^{-1}$, a P.D. matrix. Then $\mathbf{M}_{(\mathbf{x})}\mathbf{Z} \xrightarrow{p \cdot} \mathbf{Z}$.

Proof: $\mathbf{M}_{(\mathbf{x})}\mathbf{Z} = \mathbf{Z} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}/n) \xrightarrow{p \cdot} \mathbf{Z} - \mathbf{X}\mathbf{Q}^{-1}\left(E[\mathbf{X}'\mathbf{Z}]\right) = \mathbf{Z} - \mathbf{X}\mathbf{Q}^{-1}\mathbf{0} = \mathbf{Z}$ ♦

As $E[\mathbf{X}_t\delta_t] = E[\mathbf{X}_{t-1}\delta_t] = E[\mathbf{X}_t\varepsilon_t] = E[\mathbf{X}_t\varepsilon_{t-1}] = E[\mathbf{X}_{t-1}\varepsilon_t] = 0$, it follows from our lemma above the three last terms of (A2.1) have $p \cdot$ limit

$$0 + \left(\mathbf{I} - \varepsilon_{t-1}\left(\varepsilon_{t-1}'\varepsilon_{t-1}\right)^{-1}\varepsilon_{t-1}'\right)\varepsilon_{t-1}\phi + \left(\mathbf{I} - \varepsilon_{t-1}\left(\varepsilon_{t-1}'\varepsilon_{t-1}\right)^{-1}\varepsilon_{t-1}'\right)\varepsilon_{t-1}\delta_t = 0,$$

i.e. $\hat{\delta}_{\mathrm{II},i,A} - \delta_i \xrightarrow{p \cdot} 0$, hence $\hat{\tilde{\varepsilon}}_{\mathrm{II},A} \xrightarrow{\ell \cdot} \tilde{\varepsilon}$.

Appendix A3: Three-step residuals.

According to "One-step residuals" above, $\max_{1 \leq i \leq n}\left(h_{(\tilde{\mathbf{x}})ii}\right) \to 0 \Rightarrow \hat{\tilde{\varepsilon}}_{I,A} \xrightarrow{\ell \cdot} \tilde{\varepsilon}, \ \hat{\tilde{\varepsilon}}_{I,H} \xrightarrow{\ell \cdot} \tilde{\varepsilon}$.

Applying the same argument here, we need $\max_{1 \leq i \leq n}\left(h_{(\hat{\tilde{\mathbf{x}}})ii}\right) \to 0$ (i.e. we use $\hat{\mathbf{\Omega}}^{-1/2}\mathbf{X}$ rather than $\mathbf{\Omega}^{-1/2}\mathbf{X}$) for $\hat{\tilde{\varepsilon}}_{\mathrm{III},A} \xrightarrow{\ell \cdot} \tilde{\varepsilon}, \ \hat{\tilde{\varepsilon}}_{\mathrm{III},H} \xrightarrow{\ell \cdot} \tilde{\varepsilon}$ to hold. But as $\hat{\omega}_{ij} \xrightarrow{p} \omega_{ij}$, it follows that

$$\sum_{j=1}^{m} \hat{\omega}_i^{-1/2} x_{ijl}^2 \xrightarrow{p} \sum_{j=1}^{m} \omega_i^{-1/2} x_{ijl}^2 = \sum_{j=1}^{m} \tilde{x}_{ijl}^2$$ for all m sufficiently smaller than n. Hence $\max_{1 \leq i \leq n}\left(h_{(\hat{\tilde{\mathbf{x}}})}\right) \to 0$ and $\hat{\tilde{\varepsilon}}_{\mathrm{III},A} \xrightarrow{\ell \cdot} \tilde{\varepsilon}, \ \hat{\tilde{\varepsilon}}_{\mathrm{III},H} \xrightarrow{\ell \cdot} \tilde{\varepsilon}$.

References

(1) Holgersson, T. and G. Shukur, Some Aspects of Non-Normality Tests in Systems of Regression Equations. Submitted for publication in Communications, 20000.

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

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

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

(5) Kendall, M. and A. Stuart, *The Advanced Theory of Statistic, Volume 1, Distribution Theory*. 4 ed. London: Charles Griffin&Company Limited.

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

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

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

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

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

(11) Mardia, K.V., Meassures of Multivariate skewness and Kurtosis with Applications. Biometrica, 1970, *57* (3), 519-530.

(12) Dufour, J.M., and Farhat, A.; Gardiol, L.; Khalaf, L. Simulation-based finite sample normality tests in linear regressions. Econometrics Journal, 1998, *1*, 154-173.

(13) Johnson, M.E., *Multivariate Statistical Simulation*. 1987, New York: Wiley.

(14) Huber, P., *Robust Statistics*. W. 1981: Wiley.

(15) Judge, G.G.; and Griffiths, W.E.; Hill, R.C.; Lütkepohl, H.; Lee, T. *The Theory and practice of Econometrics*. 1985: Wiley.

(16) Rao, C.R., *Linear Statistical Inference*. 2:nd ed. 1973, New York: Wiley.

(17) Amemiya, T., *Advanced Econometrics*. 1 ed. 1985, Oxford: Basil Blackwell.

Research Report

1999:7	Järpe, E. & Wessman, P.:	Some power aspects of methods for detecting different shifts in the mean.
1999:8	Johnsson, T.:	On statistics and scientific thinking.
1999:9	Afsarinejad, K.:	Trend-free repeated measurement designs.
1999:10	Carlquist, A. m.fl.	The impact of season and climate on growth during early childhood in different socio--economic groups in Lahore, Pakistan.
1999:11	Carlquist, A, Erling, V. & Frisén, M.:	Longitudinal methods for analysis of the influence of breastfeeding on early child in Pakistan.
1999:12	Carlquist, A.:	Longitudinal methods for analysis of early child health in Pakistan.
1999:13	Frisén, M.:	Characterization of methods for surveillance by optimality.
2000:1	Hatemi-J, A. & Shukur, G.:	Multivariate based causality tests of twin deficits in the US.
2000:2	Petzold, M.:	Aspects on tolerance limit estimation - some common approaches and flexible modeling.
2000:3	Andersson, L.:	Statistical test of the existence of a turning point.
2000:4	Almasri, A.:	Some aspects of wavelet analysis in time series.
2000:5	Järpe, E.	Surveillance of spatio-temporal patterns. Change of interaction in an Ising dynamic model.
2000:6	Järpe, E.:	Detection of environmental catastrophes.