cemmap

centre for microdata methods and practice

Normal but Skewed?

Dante Amengual
Xinye Bei
Enrique Sentana

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP24/21

Economic and Social Research Council

Normal but Skewed?

Abstract

We propose a multivariate normality test against skew normal distributions using higher-order log-likelihood derivatives which is asymptotically equivalent to the likelihood ratio but only requires estimation under the null. Numerically, it is the supremum of the univariate skewness coefficient test over all linear combinations of the variables. We can simulate its exact finite sample distribution for any multivariate dimension and sample size. Our Monte Carlo exercises confirm its power advantages over alternative approaches. Finally, we apply it to the joint distribution of US city sizes in two consecutive censuses finding that non-normality is very clearly seen in their growth rates.

JEL Codes: C46, R11.

Keywords: City size distribution, exact test, extremum test, Gibrat's law, skew normal distribution.

Dante Amengual
CEMFI
amengual@cemfi.es

Xinyue Bei
Duke University
xinyue.bei@duke.edu

Enrique Sentana
CEMFI
sentana@cemfi.es

Acknowledgement

Some of the results in this paper appear in an earlier version of Amengual, Bei and Sentana (2021). In addition to those acknowledged there, we would like to thank audiences at Boston University, Boston College, NYU, Harvard and the World Congress of the Econometric Society for helpful comments and suggestions. We are also grateful to Jan Eeckhout and Rafael González-Val for sharing their data with us. Of course, the usual caveat applies. Financial support from the Santander - CEMFI Research Chair is gratefully acknowledged. The first and third authors also thank the Spanish Ministry of Economy, Industry & Competitiveness for its financial support through grant ECO 2017-89689.

1 Introduction

The skew-normal distribution is a generalization of the normal distribution introduced by Azzalini (1985) in the univariate case and Azzalini and Dalla Valle (1996) in the multivariate one, which allows for asymmetry and positive excess kurtosis but retains a fair amount of analytical tractability with only one additional parameters for each element of the random vector (see Azzalini and Capitanio (2014) for a thorough study of this and other closely related distributions). Among its many empirical applications across a wide range of disciplines, in economics this distribution is increasingly popular in finance and insurance, and also for stochastic frontier models (see Adcock et al (2014) and Amsler et al (2016), respectively).

However, testing normality against skew normality has been hampered by the fact that the information matrix of the unrestricted model is singular under the null of normality despite the skew normal model parameters being locally identified (see Ley and Paindaveine (2010) and Hallin and Ley (2012)). This violates one of the crucial regularity conditions that guarantees an asymptotic chi-square distribution for the Likelihood ratio (LR), Wald and score/Lagrange Multiplier (LM) tests under the null.

In the univariate case, one can overcome this problem by using the "extremum test" of Lee and Chesher (1986), which exploits the restrictions that the null imposes on higher-order optimality conditions, but which is asymptotically equivalent to the LR tests under the null and sequences of local alternatives in unrestricted contexts. Using earlier results by Cox and Hinkley (1974), Lee and Chesher (1986) explain that this equivalence intuitively follows from the fact that their extremum tests can often be re-interpreted as standard LM tests of a suitable transformation of the parameter whose first derivative is 0 on average such that the new score is no longer so. In contrast, Wald tests are extremely sensitive to reparametrization under these circumstances.[1]

In the multivariate case, though, the information matrix of the skew normal is repeatedly singular, in the sense that its nullity coincides with the dimension of random vector K. In addition, there are K linear combinations involving the elements of the score vector and the Hessian matrix which are also 0 under the null, which means that it is necessary to look at third-order derivatives. Unfortunately, the number of such derivatives exceeds the number of parameters effectively affected by the singularity by two orders of magnitude, so there is no natural reparametrization leading to a regular information matrix. In particular, transforming each of the parameters individually along the lines suggested by Lee and Chesher (1986) does not give rise to a test asymptotically equivalent to the LR. On the contrary, different reparametrizations will typically give rise to different test statistics.

The purpose of our paper is to derive a test of multivariate normality against skew normal

[1] Rotnitzky el al (2000) rigorously study the asymptotic distribution of the maximum likelihood (ML) estimators when there is a single singularity, while Bottai (2003) looks at the validity of confidence intervals obtained by inverting the three classical test statistics in the same setup.

distributions which is asymptotically equivalent to the LR test, but which only requires estimation under the null. To do so, we rely on the generalized extremum tests we proposed in a companion paper (see Amengual, Bei and Sentana (2021)). As we show below, the resulting test statistic has a very simple interpretation in terms of moment tests. Specifically, it numerically coincides with the supremum of the tests for univariate asymmetry based on the sample skewness coefficient over all possible linear combinations of the observed variables.

Importantly, we explicitly address the widespread and often justified concern that tests based on higher-order derivatives are unreliable in finite samples by explaining how to simulate its exact, parameter-free, finite sample distribution to any desired degree of accuracy for any dimension of the random vector and sample size. In this respect, we prove the numerical invariance of the test statistic to affine transformations of the observed variables, which allows us to quickly simulate draws from a spherical normal distribution.

We conduct extensive Monte Carlo exercises that study the finite sample size and power properties of our proposal and compare it to other multivariate skewness tests. We find that our suggested parametric bootstrap procedure yields very reliable sizes. In addition, we confirm the power superiority of our test over the alternatives. We also confirm its computational superiority over the LR test, which is due to the fact that it does not require the estimation of any additional shape parameters using a log-likelihood function which is incredibly flat under the null.

Finally, we illustrate our test by looking at the joint distribution of city sizes in the 2000 and 2010 US censuses. The starting point of our empirical analysis is Eeckhout (2004), who forcefully argued that if one looked at the entire non-truncated sample of cities and places in the 2000 US census, their size distribution was approximately log-normal, in marked contrast to earlier studies. Subsequent studies have analyzed the same issue for other datasets for the US and other countries (see e.g. Gónzalez-Val (2019) and the references therein), but they have not typically looked at the joint distribution of city sizes in two periods. An important advantage of looking at two censuses is that we can immediately study the joint distribution of city sizes and their rates of growth. In this regard, a useful shared property of multivariate normality and multivariate skew normality is that they are both closed under affine transformations of the original variables (see Azzalini and Capitanio (2014)). Importantly, we find that skewness is a common feature that is much more clearly seen in the growth rate of cities than in their (log) sizes.

The rest of the paper is organized as follows. In Section 2, we derive our proposed test and study its properties. This is followed by the simulation exercises in Section 3, and the empirical application in Section 4. Finally, we present our conclusions in section 5. Proofs and auxiliary results are gathered in appendices.

2 Testing Gaussian vs Skew Normal

The probability density function (pdf) of a K-dimensional skew-normal random variable $\mathbf{y}$ is given by

$$f_{SN}(\mathbf{y};\boldsymbol{\varrho}) = 2f_N(\mathbf{y};\boldsymbol{\varphi}_M,\boldsymbol{\varphi}_V)\cdot\Phi[\boldsymbol{\vartheta}'\mathrm{dg}^{-1/2}(\boldsymbol{\varphi}_D)(\mathbf{y}-\boldsymbol{\varphi}_M)], \quad (1)$$

where $f_N(\boldsymbol{y};\boldsymbol{\varphi}_M,\boldsymbol{\varphi}_V)$ denotes the pdf of a K-variate Gaussian random vector with mean $\boldsymbol{\varphi}_M$ and covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)$, $\boldsymbol{\varphi}_V = (\boldsymbol{\varphi}_D', \boldsymbol{\varphi}_L')'$, $\boldsymbol{\varphi}_D = vecd[\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)]$, $\boldsymbol{\varphi}_L = vecl[\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)]$, $\boldsymbol{\varrho}' = (\boldsymbol{\varphi}', \boldsymbol{\vartheta}') = (\boldsymbol{\varphi}_M', \boldsymbol{\varphi}_V', \boldsymbol{\vartheta}')$, $\mathrm{dg}(\boldsymbol{\varphi}_D)$ a diagonal matrix with $\boldsymbol{\varphi}_D$ along its main diagonal, and $\Phi(.)$ the univariate standard normal cumulative distribution function (cdf). This joint distribution simplifies to the K-variate normal when the shape parameters $\boldsymbol{\vartheta}$ are equal to $\mathbf{0}$.

For expositional purposes, we use the bivariate case, which is the relevant one for our empirical application. Nevertheless, our theoretical results apply to any K.

Let $\boldsymbol{\varphi} = (\varphi_{M_1}, \varphi_{M_2}, \varphi_{D_1}, \varphi_{D_2}, \varphi_{L_1})'$ and $\boldsymbol{\vartheta} = (\vartheta_1, \vartheta_2)'$ denote the vectors that contain the two mean and three covariance parameters, and the two shape parameters, respectively, so that

$$f_{SN}(\mathbf{y};\boldsymbol{\varphi},\boldsymbol{\vartheta}) = 2f_N\left[\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}; \begin{pmatrix} \varphi_{M_1} \\ \varphi_{M_2} \end{pmatrix}, \begin{pmatrix} \varphi_{D_1} & \varphi_{L_1} \\ \varphi_{L_1} & \varphi_{D_2} \end{pmatrix}\right]\cdot\Phi\left[\vartheta_1\left(\frac{y_1-\varphi_{M_1}}{\sqrt{\varphi_{D_1}}}\right)+\vartheta_2\left(\frac{y_2-\varphi_{M_2}}{\sqrt{\varphi_{D_2}}}\right)\right].$$

As explained by Azzalini and Capitanio (2014), the information matrix of this model is generally regular, so that the unrestricted Maximum Likelihood estimators (MLEs) of $\boldsymbol{\varphi}_M$, $\boldsymbol{\varphi}_V$ and $\boldsymbol{\vartheta}$ based on a random sample of $\mathbf{y}$ will have an asymptotic normal distribution. In addition, the restricted MLEs of $\boldsymbol{\varphi}_M$ and $\boldsymbol{\varphi}_V$ under the null of multivariate normality will coincide with the sample mean vector and covariance matrix of the observations (with denominator the sample size n), which also have a well-known asymptotic normal distribution under the null.

Nevertheless, it is easy to prove that when evaluated at $\boldsymbol{\vartheta} = \mathbf{0}$, the score of each element of $\boldsymbol{\vartheta}$ is proportional to the score of the corresponding element $\boldsymbol{\varphi}_M$ regardless of the sample size, which confirms the repeated singularity of the information matrix of the model under the null. In addition, it is also easy to prove that $K(K+1)/2$ additional independent linear combinations of the elements of the Hessian matrix and the score vector are also 0 when $\boldsymbol{\vartheta} = \mathbf{0}$. As a result, the joint asymptotic distribution of the unrestricted MLEs of $\boldsymbol{\varphi}_M$, $\boldsymbol{\varphi}_V$ and $\boldsymbol{\vartheta}$ will not be normal when the true distribution is normal, and the LR test will not have a chi-square distribution either. In addition, obtaining the distribution of the LR test under the null by simulation is an extremely challenging procedure from a computational point of view because for each simulated sample it requires the maximization with respect to all the elements of $\boldsymbol{\varphi}_M$, $\boldsymbol{\varphi}_V$ and $\boldsymbol{\vartheta}$ of a log-likelihood function which is extremely flat along those directions of the parameter space whose first and second derivatives are 0.

In this context, we can state our main result:

Proposition 1 *The difference between LR test of $H_0 : \boldsymbol{\vartheta} = \mathbf{0}$ in model (1) based on a random* *sample of n observations on $\mathbf{y}$ and the following test statistic*

$$GET_n = \sup_{\boldsymbol{\lambda}\neq\mathbf{0}} \frac{1}{6n}\left[\sum_{i=1}^{n} H_3\left(\frac{\boldsymbol{\lambda}'\mathbf{e}_i}{\sqrt{\boldsymbol{\lambda}'\boldsymbol{\lambda}}}\right)\right]^2 \tag{2}$$

is $O_p(n^{-1/6})$, where $H_3(z) = z^3 - 3z$ is the third-order Hermite polynomial of a standardized variable z, $\boldsymbol{\lambda}$ is a real vector of dimension K and $\mathbf{e}$ denotes an affine transformation of the observed variables whose sample mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_K$, respectively.

In simple terms, our test statistics numerically coincides with the supremum of the moment tests for univariate skewness based on the third Hermite polynomial over all possible linear combinations of the observed variables that have 0 mean and unit variance in the sample. In fact, the standardization is unnecessary because the moment test for univariate skewness is numerically invariant to affine transformations of the observations, which in turn confirms that the test statistic (2) is homogeneous of degree 0 in $\boldsymbol{\lambda}$. Thus, when $K = 1$ our proposed test reduces to the well known moment test for univariate skewness based on the third Hermite polynomial of the standardized observations. This test of normality versus skew normality, which Carota (2010) derived using a divergence-based Bayesian method, can be obtained as a straightforward application of the Lee and Chesher (1986) extremum test by replacing the skewness coefficient ϑ by its cubic root.

As we shall formally show in Proposition 2 below, the particular transformation that orthonormalizes the variables in the sample is irrelevant. For example, in the bivariate case, we could define e_1 as the standardized value y_1 and e_2 as the standardized value of the residual in the OLS regression of y_2 on a constant and y_1. But we could also define them the other way round.

The bivariate case provides some further insight. Given that the sample means and variances of e_1 and e_2 are 0 and 1, respectively, we can write the test statistic as

$$\mathrm{GET}_n = \sup_{\|\lambda\|=1} \frac{1}{6n}\left[\begin{array}{c} \lambda_1^3\sum_{i=1}^n H_3(e_{1i}) + 3\lambda_1^2\lambda_2\sum_{i=1}^n H_2(e_{1i})H_1(e_{2i}) \\ +3\lambda_1\lambda_2^2\sum_{i=1}^n H_1(e_{1i})H_2(e_{2i}) + \lambda_2^3\sum_{i=1}^n H_3(e_{2i}) \end{array}\right]^2$$

where $H_1(z) = z$ and $H_2(z) = z^2 - 1$ are the first- and second-order Hermite polynomials of the standardized variable z. Therefore, the first and last of the four terms of the test statistic effectively check the asymmetry of the marginal distributions of e_1 and e_2 by looking at their third-order Hermite polynomials $H_3(e_1)$ and $H_3(e_2)$, respectively, while the two middle ones check the co-asymmetries between those two random variables by focusing on $H_2(e_1)H_1(e_2)$ and $H_1(e_1)H_2(e_2)$.

Consider now the following full-rank affine transformation $\mathbf{y}^* = \mathbf{a} + \mathbf{B}\mathbf{y}$ with $|\mathbf{B}| \neq 0$. A convenient property of the skew normal distribution that it shares with its Gaussian special case is that it is closed under affine transformations (see Azzalini and Capitanio (2014)). Thus, $\mathbf{y}^*$ will be normal if and only if $\mathbf{y}$ is normal, but it will be skew normal otherwise. Our next result shows that the test statistic in Proposition 1 is numerically invariant to the values of $\mathbf{a}$ and $\mathbf{B}$:

Proposition 2 *The generalized extremum test statistic of model (1) numerically coincides with the analogous test statistic for* $\mathbf{y}^*$.

This numerical invariance is not only a very desirable property of any multivariate normality test, as forcefully argued by Henze (2002), but it also implies that the sample mean vector and covariance matrix of the observations do not affect the null distribution of our proposed test in finite samples. As a result, it is possible to simulate its exact, parameter-free, finite sample distribution to any desired degree of accuracy for any dimension of the random vector and sample size. In particular, it suffices to simulate R times a random sample of size n of a spherical Gaussian random vector of dimension K to obtain R independent draws of the test statistic for multivariate normality against a skew normal. This can be regarded as a parametric bootstrap procedure that provides the exact p-value of the test statistic obtained in a real sample as the number of bootstrap replications R grows without bound. But the fact that the only characteristics of the original sample that matter are the values of n and K implies that a researcher could obtain tables with exact critical values before observing the data.

It is worth emphasizing that the maximization required to compute (2) is over $K-1$ dimensions only, as opposed to the maximization required for the computation of the LR test, which is effectively over $2K$ parameters because $\boldsymbol{\varphi}_V$ can be concentrated out (see Azzalini and Capitanio (2014, sec. 5.2.1)). For example, in the bivariate case, if we express $\lambda_1 = \cos\omega$ and $\lambda_2 = \sin\omega$, it simply requires finding the optimal angle ω over $(0, \pi)$, which can be done very accurately in very little time.

3 Simulation evidence

In this section, we study the finite sample size and power properties of the test we have introduced in Proposition 1 by means of some extensive Monte Carlo exercises, comparing it to other skewness tests. Specifically, for each of the distributional assumptions we describe below, we generate 10,000 samples of size $n = 400$ and $n = 1,600$ and in each of them we compute (2) together with the following three alternative testing procedures:

1) A joint test that simultaneously looks at the moment conditions $E\{H_3[\varphi_{D,k}^{-1/2}(y_k - \varphi_{M,k})]\} = 0$ for $k = 1, \ldots, K$, where y_k is the k^{th} element of $\mathbf{y}$.

2) A joint test that simultaneously looks at the moment conditions $E[H_{klm}(\mathbf{y}; \boldsymbol{\varphi}_M, \boldsymbol{\varphi}_V)] = 0$ for all the $K(K+1)(K+2)/6$ different third-order multivariate Hermite polynomials of the form

$$H_{klm}(\mathbf{y}; \boldsymbol{\varphi}_M, \boldsymbol{\varphi}_V) = -e^{\frac{1}{2}(\mathbf{y}-\boldsymbol{\varphi}_M)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)(\mathbf{y}-\boldsymbol{\varphi}_M)} \frac{\partial^3}{\partial y_k \partial y_l \partial y_m} \left[e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\varphi}_M)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)(\mathbf{y}-\boldsymbol{\varphi}_M)} \right], \quad (3)$$

with $k, l, m = 1, \ldots, K$.

3) A joint test that simultaneously looks at the moment conditions $E[H_{kkk}(\mathbf{y}; \boldsymbol{\varphi}_M \boldsymbol{\varphi}_V)] = 0$ for $k = 1, \ldots, K$.

The first test, which is simply looking at the marginal skewness of the observed variables, ignores all the $K(K-1)(K+4)/6$ co-skewness terms. As a result, the value of the test statistic changes when we consider affine transformations of the observations. More importantly, its power will crucially depend on the marginal skewness of the transformed variables, so it cannot be asymptotically equivalent to the LR test, which is numerically and therefore asymptotically invariant to affine transformations because both normal and skew normal distributions are closed under such transformations.

Similarly, the third test, which we derive in Appendix B by applying the Lee and Chesher (1986) approach to each of the shape parameters of the central parametrization proposed by Arellano-Valle and Azzalini (2008), is not numerically invariant to affine transformations either because those shape parameters also capture the marginal skewness of the transformed variables. For that reason, this test cannot be asymptotically equivalent to the LR test.[2]

In contrast, the second test, which coincides with the skewness component of Mardia's (1970) test for multivariate normality, is numerically invariant to affine transformations of the observations. Unfortunately, this test fails to exploit that for multivariate skew normal distributions skewness is a common feature (see Engle and Kozicki (1993)). Specifically, Theorem 5.12 in Azzalini and Capitanio (2014) states that there is always a linear "canonical" transformation of the observed variables in which one marginal distribution is univariate skew normal but the other $N-1$ variables are Gaussian and stochastically independent. Given that all the remaining third and fourth multivariate cumulants are 0, this representation implies that the only effect of increasing K is to add more independent Gaussian components, which in turn add more 0 (co-)skewness and (co-)kurtosis terms. As a result, the non-centrality parameter of the second test remains the same as K grows, while the number of degrees of freedom increases, which results in a loss of power.

The main advantage of these three tests is that their asymptotic distribution under the null hypothesis is chi-square with as many degrees of freedom as moments involved. In contrast, the test in Proposition 1 has a non-standard asymptotic distribution, which it shares with the LR test. In principle, we could bound this asymptotic distribution from below by the univariate skewness test of any linear combination of the observed series, including the margins, which converges to a χ^2_1 for fixed $\boldsymbol{\lambda}$. Similarly, we could bound it from above by the skewness component of Mardia's (1970) multivariate normality test, which converges to a $\chi^2_{K(K+1)(K+2)/6}$. However, those bounds become increasingly loose as K increases. As we explained in the previous section, though, in practice we can easily compute by simulation very good approximations to the exact critical values under the null for any n and K.

As alternative hypotheses, we keep $\boldsymbol{\varphi}_M = \mathbf{0}$ and $\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V) = \mathbf{I}_N$ but consider

$$\boldsymbol{\vartheta}' = \left(\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right) \quad (H_{a1}) \quad \text{and} \quad \boldsymbol{\vartheta}' = \left(\sqrt{\frac{3}{10}}, 2\sqrt{\frac{3}{10}}\right) \quad (H_{a2})$$

[2] We can prove that $H_{kkk}(\mathbf{y}; \boldsymbol{\varphi}_M, \boldsymbol{\varphi}_V) \propto H_3(z_k^*)$, where z_k^* is the residual in the theoretical regression of y_k on a constant and the remaining elements of $\mathbf{y}$ divided by the standard error from that regression, so that the tests in 1) and 3) will only be asymptotically equivalent under the null when the original variables are orthogonal to each other, as in our simulations.

in the bivariate case, and

$$\boldsymbol{\vartheta}' = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \quad (H_{a1}) \quad \text{and} \quad \boldsymbol{\vartheta}' = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right) \quad (H_{a2})$$

in the trivariate one. All these distributions are such that the skewness coefficient of the linear combination of the observed variables that absorbs all the skewness in the canonical representation is $\sqrt{3/2}$. Consequently, the power of our proposed test should be the same for each combination of n and K regardless of the value of $\boldsymbol{\vartheta}$. The same is true of Mardia's (1970) test but not of the other two.

Panels A and B of Table 1 report the results for bivariate and trivariate models, respectively. The first three columns contain rejection rates under the null at the 1%, 5% and 10% levels, confirming that our simulated critical values work remarkably well for both sample sizes.[3]

In turn, the last six columns present the rejection rates at the 1%, 5% and 10% levels for the alternatives we consider. The first thing we observe is that the powers of both GET and Mardia's (1970) tests are effectively identical for H_{a1} and H_{a2} for each combination of n and K, unlike what happens to the other two tests.

More importantly, our proposed test is more powerful than the third test above for all alternatives. It also beats by far the test based on the skewness coefficients of the margins. Interestingly, the skewness component of Mardia's (1970) test has similar power to ours in the bivariate case, but it losses power in the trivariate case, as expected from our previous discussion.

Finally, Figure 1a visually confirms that the Gaussian rank correlation coefficients[4] between the GET and LR test statistics across Monte Carlo samples of size $n = 400$ and 1,600 generated under the null are .97 and .98, respectively. The same pattern is also present in the trivariate case depicted in Figure 1b, which is in line with the asymptotic equivalence result in Proposition 1. Our results also indicate that the LR takes between 12 and 75 times as much CPU time to compute as GET does. More importantly, we often find that the log-likelihood function has many flat areas under the alternative, which leads the numerical optimization algorithm to stop at a value below the maximum. In this respect, we find that using as starting value for the skewness direction the $\boldsymbol{\lambda}$ that maximizes GET_n helps.

4 The distribution US city sizes and their growth rates

As we mentioned in the introduction, Eeckhout (2004) forcefully argued that if one looked at the entire untruncated sample of cities and places in the 2000 US census, their size distribution was approximately log-normal, in marked contrast to earlier studies. Subsequent papers have

[3] With 10,000 Monte Carlo replications, the 95% asymptotic confidence intervals for the Monte Carlo rejection probabilities under the null are (.80,1.20), (4.57,5.43) and (9.41,10.59) at the 1, 5 and 10% levels.

[4] The Gaussian rank correlation coefficient between two variables is the usual Pearson correlation coefficient between the Gaussian scores of those variables, which are obtained by applying the inverse Gaussian cdf transform to the ranks of the observations on each variable divided by $n+1$ (see Amengual, Sentana and Tian (2020)). Like the Spearman correlation coefficient, it is is less sensitive to outliers than the Pearson one.

analyzed the same issue for other datasets from the US and other countries (see e.g. González-Val (2019) and the references therein), but they have not typically looked at the joint distribution of city sizes at two different points in time. An important advantage of looking at two censuses is that we can simultaneously study the joint distribution of initial (log) city size and its (geometric) rate of growth, whose independence is at the core of Gibrat's law.

We follow the extant literature and treat Alaska, Hawaii and the remaining off-shore insular territories like Puerto Rico separately from the remaining contiguous 48 states in the North American continent. Changes in boundaries and city names, as well as the creation of new entities and the dissolution of others, imply that there is no one-to-one relationship between the entity names and codes in the 2000 and 2010 censuses files. In practice, though, the differences are small. Specifically, in the 2010 census we can match 24,023 of the 24,670 places that appeared in the 2000 census. For that reason, we follow Eeckhout (2004) and look at the joint distribution of the 24,009 matched cities with a population of at least one in both years.

Figures 2a and 2b contain kernel density estimates of the marginal distributions of (log) city sizes for the contiguous states in 2000 and 2010, together with the best normal approximations to them, which share their sample means and standard deviations. Although both estimated densities differ from their normal approximations, at first sight there is not much evidence of kurtosis and only some evidence of asymmetries around the mode of the distributions rather than at the tails. Standard normality tests for univariate distributions confirm both these impressions. Specifically, the kurtosis of the marginal distributions in the two periods is 3.03 and 2.98, which are not statistically significantly different from 3. As acknowledged by Eeckhout (2004) for the 2000 data, though, the skewness coefficients are positive (0.25 and 0.21, respectively) and statistically significant in view of the large number of observations.

Our main interest, though, is the bivariate distribution. Figure 3 contains a scatter plot of (log) city sizes for the contiguous states in 2000 and 2010, together with level curves for the corresponding bivariate kernel density estimate. Clearly, the joint distribution seems far more non-normal than any of the margins. This is confirmed by our GET test, which is equal to 1,772,758 with a negligible exact p-value.

Interestingly, we can reverse engineer the fact that our test statistic is the supremum of the moment tests for skewness of all possible linear combinations of the two log-sizes to find out which linear combination is the most non-normal. Somewhat surprisingly, we find that the transformation of the original variables that maximizes the statistic in the sample has weights (proportional to) $(-1.04, 1)$, which means that skewness is a feature that is much more clearly seen in the growth rate of cities than in their (log) sizes.

We confirm this finding by looking at the distribution of growth rates between 2000 and 2010 in Figure 4, which is not only far more peaked (kurtosis = 45.09) but also substantially more asymmetric (skewness = 0.32). We can obtain a complementary perspective on the joint distribution by looking at the joint distribution of (log) city sizes in 2000 and (geometric) growth rates between 2000 and 2010. Although the scatter plot in Figure 5 is a simple linear transformation of the one in Figure 3 in which the previous 45^o degree line has become the new vertical axis, it arguably makes the non-normality of the joint distribution far more evident. Our test statistic, though, is invariant to this transformation.

5 Conclusions

In this paper, we have developed a multivariate normality test against skew normal distributions which is asymptotically equivalent to the LR but only requires estimation under the null. It overcomes the singularities of the elements of the score vector and Hessian matrix associated to the shape parameters by working with third-order derivatives. Importantly, we prove that it coincides with the supremum of the univariate skewness coefficient test over all linear combinations of the variables. We also explain how to simulate its exact finite sample distribution for any dimension of the random vector and sample size. Our Monte Carlo exercises confirm its power advantages over alternative approaches and its computational advantages over the LR. When we apply it to the joint distribution of US city sizes in two consecutive censuses, we find that non-normality is very clearly seen in their growth rates.

From the theoretical point of view, the development of tests of multivariate normality against multivariate skewed t distributions provides an interesting avenue for additional research. A more thorough study of the potential dependence of the distribution of growth rates and initial city size also deserves further investigation. Similarly, we could study the possibility of using the test in Proposition 1 to obtain uniform confidence intervals, as in Bottai (2003).

References

Adcock, C., Eling, M. and Loperfido, N. (2015): "Skewed distributions in finance and actuarial science: a review", *The European Journal of Finance* 21, 1253–1281.

Amengual, D., Bei, X. and Sentana, E. (2021): "Hypothesis tests with a repeatedly singular information matrix", mimeo, CEMFI.

Amengual, D., Sentana, E. and Tian, Z. (2020): "Gaussian rank correlation and regression", forthcoming in A. Chudik, C. Hsiao and A. Timmermann (eds.) *Essays in honor of M. Hashem Pesaran*, *Advances in Econometrics*, Emerald.

Amsler, C., Prokhorov, A. and Schmidt, P. (2016): "Endogeneity in stochastic frontier models", *Journal of Econometrics* 190, 280-288.

Arellano-Valle, R.B. and Azzalini, A. (2008): "The centred parametrization for the multivariate skew-normal distribution", *Journal of Multivariate Analysis* 99, 1362–1382.

Azzalini, A. (1985): "A class of distributions which includes the normal ones", *Scandinavian Journal of Statistics* 12, 171–178.

Azzalini, A. and Capitanio, A. (2014): *The skew-normal and related families*, IMS monographs. Cambridge University Press.

Azzalini, A. and Dalla Valle, A. (1996): "The multivariate skew-normal distribution", *Biometrika* 83, 715–726.

Bottai, M. (2003): "Confidence regions when the Fisher information is zero", *Biometrika* 90, 73–84.

Carota, C. (2010): "Tests for normality in classes of skew-t alternatives", *Statistics and Probability Letters* 80, 1–8.

Cox, D. and Hinkley, D. (1974): *Theoretical statistics*, Chapman and Hall.

Eeckhout, J. (2004): "Gibrat's law for (all) cities", *American Economic Review* 94, 1429–1451.

Engle, R.F. and Kozicki, S. (1993): "Testing for common features", *Journal of Business and Economic Statistics* 11, 369–380.

González-Val, R. (2019): "US city-size distribution and space", *Spatial Economic Analysis* 14, 283-300.

Lee, L. F. and A. Chesher (1986): "Specification testing when score test statistics are identically zero", *Journal of Econometrics* 31, 121–149.

Ley, C. and Paindaveine, D. (2010): "On the singularity of multivariate skew-symmetric models", *Journal of Multivariate Analysis* 101,1434–1444.

Hallin, M. and Ley, C. (2012): "Skew-symmetric distributions and Fisher information – a tale of two densities", *Bernoulli* 18, 747-763.

Mardia, K.V. (1970): "Measures of multivariate skewness and kurtosis with applications", *Biometrika* 57, 519-530.

Henze, N. (2002): "Invariant tests for multivariate normality: a critical review", *Statistical Papers* 43, 467–506.

Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000): "Likelihood-based inference with singular information matrix", *Bernoulli* 6, 243–284.

Appendices

A Proofs

A.1 Proof of Proposition 1

Given the density of a multivariate skew normal random vector in (1), the contribution of $\mathbf{y}$ to the log-likelihood function is

$$l(\mathbf{y};\boldsymbol{\varrho}) = \text{const} + T_1(\mathbf{y};\boldsymbol{\varrho}) + T_2(\mathbf{y};\boldsymbol{\varrho}) + T_3(\mathbf{y};\boldsymbol{\varrho}), \tag{A1}$$

where $\boldsymbol{\varrho} = (\boldsymbol{\varphi}_M', \boldsymbol{\varphi}_V', \boldsymbol{\vartheta}')'$,

$$T_1(\mathbf{y};\boldsymbol{\varrho}) = -\frac{1}{2}\log\{\det[\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)]\}, \tag{A2}$$

$$T_2(\mathbf{y};\boldsymbol{\varrho}) = -\frac{1}{2}(\mathbf{y}-\boldsymbol{\varphi}_M)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)(\mathbf{y}-\boldsymbol{\varphi}_M), \tag{A3}$$

$$T_3(\mathbf{y};\boldsymbol{\varrho}) = \log\{\Phi[\boldsymbol{\vartheta}'\text{dg}^{-\frac{1}{2}}(\boldsymbol{\varphi}_D)(\mathbf{y}-\boldsymbol{\varphi}_M)]\}. \tag{A4}$$

Since the information matrix is repeatedly singular, the first thing we do is to express the model in terms of an alternative set of parameters $\boldsymbol{\rho} = (\boldsymbol{\phi}_M', \boldsymbol{\phi}_V', \boldsymbol{\theta}')'$ such that:

$$\begin{aligned}
\boldsymbol{\varphi}_M &= \boldsymbol{\phi}_M - \sqrt{\frac{2}{\pi}}\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\text{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\theta} \\
\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V) &= \boldsymbol{\Omega}(\boldsymbol{\phi}_V) + \frac{2}{\pi}\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\text{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\theta}\boldsymbol{\theta}'\text{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V) \\
\boldsymbol{\vartheta} &= \boldsymbol{\theta}
\end{aligned}$$

where $\boldsymbol{\Omega}(\boldsymbol{\phi}_V)$ is a $K\times K$ symmetric matrix such that $vecd[\boldsymbol{\Omega}(\boldsymbol{\phi}_V)] = \boldsymbol{\phi}_D$ and $vecl[\boldsymbol{\Omega}(\boldsymbol{\phi}_V)] = \boldsymbol{\phi}_L$, with $\boldsymbol{\phi}_V = (\boldsymbol{\phi}_D', \boldsymbol{\phi}_L')'$. This one-to-one reparametrization simply concentrates the singularity of the information matrix on the new skewness parameters so that all the elements of the score vector and Hessian matrix corresponding to $\boldsymbol{\theta}$ are zero at once.

We skip the verification of the regularity conditions in Assumption 1 in Amengual, Bei and Sentana (2021) because they are straightforward. To check that their Assumption 2 holds for $\boldsymbol{\rho}$, in what follows we avoid the complex notation necessary for higher-order matrix derivatives by letting $\boldsymbol{\theta} = \boldsymbol{\lambda}\eta$ for a fixed arbitrary vector $\boldsymbol{\lambda}\in\mathbb{R}^K$, $\boldsymbol{\lambda}\neq\mathbf{0}$, so that we can simply take higher order derivatives with respect to the scalar η. Intuitively, the reason is that

$$\boldsymbol{\lambda}^{\otimes r\prime}\frac{\partial^r l}{(\partial\boldsymbol{\theta})^{\otimes r}} = \frac{\text{d}^r l}{\text{d}\eta^r},$$

where $\lambda^{\otimes r} = \underbrace{\boldsymbol{\lambda}\otimes .\overset{r}{.}. \otimes\boldsymbol{\lambda}}$ denotes the k^{th} Kronecker power of the vector $\boldsymbol{\lambda}$, and

$$\frac{\partial^k l}{(\partial\boldsymbol{\theta})^{\otimes k}} = vec\left\{\frac{\partial}{\partial\boldsymbol{\theta}}\left[\frac{\partial^{k-1} l}{\partial\boldsymbol{\theta}^{\otimes(k-1)}}\right]'\right\}.$$

In this notation, Assumption 2.1 in Amengual, Bei and Sentana (2021) is equivalent to

$$\boldsymbol{\lambda}'\frac{\partial l}{\partial\boldsymbol{\theta}} = 0, \quad \boldsymbol{\lambda}^{\otimes 2\prime}\frac{\partial^2 l}{(\partial\boldsymbol{\theta})^{\otimes 2}} = 0 \ \forall\boldsymbol{\lambda}, \tag{A5}$$

while Assumption 2.2 is equivalent to the matrix

$$\left(\frac{\partial l}{\partial\boldsymbol{\varphi}_M}, \frac{\partial l}{\partial\boldsymbol{\varphi}_V}, \boldsymbol{\lambda}^{\otimes\prime 3}\frac{\partial l}{(\partial\boldsymbol{\theta})^{\otimes 3}}\right) \tag{A6}$$

having full rank $\forall \boldsymbol{\lambda} \neq 0$.

The first derivative of (A2) with respect to η is

$$\begin{aligned}\frac{\mathrm{d}T_1(\mathbf{y};\boldsymbol{\varrho})}{\mathrm{d}\eta} &= -\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)\frac{\mathrm{d}\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)}{\mathrm{d}\eta}\right] \\ &= -\frac{2}{\pi}\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)]\eta. \end{aligned} \tag{A7}$$

Thus, we have that

$$\frac{\mathrm{d}T_1(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta} = 0 \tag{A8}$$

because (A7) is linear in η, where by $\mathrm{d}f(\boldsymbol{\varrho}_0)/\mathrm{d}\eta$ we mean the partial derivative of the function $f(\boldsymbol{\varrho})$ with respect to η evaluated at $\boldsymbol{\varrho} = \boldsymbol{\varrho}_0$.

If we derive (A7) with respect to η again, we obtain

$$\begin{aligned}\frac{\mathrm{d}^2T_1(\mathbf{y};\boldsymbol{\varrho})}{(\mathrm{d}\eta)^2} =& -\frac{2}{\pi}\mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)] \\ &- \frac{2}{\pi}\frac{\mathrm{d}\ \mathrm{tr}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)]}{\mathrm{d}\eta}\eta. \end{aligned} \tag{A9}$$

Letting

$$\boldsymbol{\Psi}(\boldsymbol{\phi}_V) = \mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D), \tag{A10}$$

we have that

$$\frac{\mathrm{d}^2T_1(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2} = -\frac{2}{\pi}\mathrm{tr}[\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)] = -\frac{2}{\pi}\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}, \tag{A11}$$

where in the first equality we have used the fact that $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)\boldsymbol{\Omega}(\boldsymbol{\phi}_V) = \mathbf{I}_K$ under the Gaussian null, while the second equality follows from the invariance of the trace of a matrix product to cyclic permutations of the factors.

We can also show that the derivative of (A9) with respect to η evaluated under the Gaussian null will be

$$\frac{\mathrm{d}^3T_1(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3} = 0. \tag{A12}$$

Let

$$\boldsymbol{\Xi}(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y}) = \boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\Sigma}\left(\boldsymbol{\varphi}_V\right)^{-1}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right).$$

We can easily prove that the first derivative of (A3) with respect to η is

$$\begin{aligned}\frac{\mathrm{d}T_2(\mathbf{y};\boldsymbol{\varrho})}{\mathrm{d}\eta} =& \frac{\mathrm{d}\boldsymbol{\varphi}_M'}{\mathrm{d}\eta}\boldsymbol{\Sigma}\left(\boldsymbol{\varphi}_V\right)^{-1}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right) - \frac{1}{2}(\mathbf{y}-\boldsymbol{\varphi}_M)'\frac{\mathrm{d}\boldsymbol{\Sigma}\left(\boldsymbol{\varphi}_V\right)^{-1}}{\mathrm{d}\eta}(\mathbf{y}-\boldsymbol{\varphi}_M) \\ =& -\sqrt{\frac{2}{\pi}}\boldsymbol{\Xi}(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y}) + \frac{2}{\pi}\boldsymbol{\Xi}'(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y})\boldsymbol{\Xi}(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y})\eta, \end{aligned} \tag{A13}$$

where the second line follows from the fact that

$$\frac{\mathrm{d}\boldsymbol{\varphi}_M'}{\mathrm{d}\eta} = -\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)$$

and

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\varphi}_V\right)}{\mathrm{d}\eta} &= -\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\varphi}_V\right)\frac{\mathrm{d}\boldsymbol{\Sigma}\left(\boldsymbol{\varphi}_V\right)}{\mathrm{d}\eta}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\varphi}_V\right) \\ &= -\frac{4}{\pi}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\varphi}_V\right)\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\lambda}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\varphi}_V\right)\eta. \end{aligned}$$

Under the Gaussian null, $\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)=\mathbf{I}_K$ and the linear term in η vanishes, so that

$$\frac{\mathrm{d}T_2(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}=-\sqrt{\frac{2}{\pi}}[\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}]^{\frac{1}{2}}Z_{\tilde{\boldsymbol{\lambda}}}, \tag{A14}$$

where

$$\begin{aligned} Z_{\tilde{\boldsymbol{\lambda}}} &\equiv [\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}]^{-\frac{1}{2}}\tilde{\boldsymbol{\lambda}}'\mathbf{Z}, \\ \tilde{\boldsymbol{\lambda}} &= \boldsymbol{\Psi}^{\frac{1}{2}}(\boldsymbol{\phi}_V)\boldsymbol{\lambda} \end{aligned} \tag{A15}$$

and

$$\mathbf{Z}=\boldsymbol{\Psi}^{-\frac{1}{2}}(\boldsymbol{\phi}_V)\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)(\mathbf{y}-\boldsymbol{\varphi}_M)$$

in view of (A10). Importantly, the choice of square root matrix $\boldsymbol{\Psi}^{\frac{1}{2}}(\boldsymbol{\phi}_V)$ is irrelevant.

The derivative of (A13) with respect to η yields

$$\begin{aligned} \frac{\mathrm{d}^2T_2(\mathbf{y};\boldsymbol{\varrho})}{(\mathrm{d}\eta)^2}= &-\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\frac{\mathrm{d}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)(\mathbf{y}-\boldsymbol{\varphi}_M)\right]}{\mathrm{d}\eta} \\ &+\frac{4}{\pi}\boldsymbol{\Xi}'(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y})\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\frac{\mathrm{d}\left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)(\mathbf{y}-\boldsymbol{\varphi}_M)\right]}{\mathrm{d}\eta}\eta \\ &+\frac{2}{\pi}\boldsymbol{\Xi}'(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y})\boldsymbol{\Xi}(\boldsymbol{\varrho},\boldsymbol{\phi},\boldsymbol{\lambda},\mathbf{y}). \end{aligned} \tag{A16}$$

Since the linear term in η vanishes under the Gaussian null, using (A10) and (A15) we obtain

$$\frac{\mathrm{d}^2T_2(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}=-\frac{2}{\pi}\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}+\frac{2}{\pi}\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}Z^2_{\tilde{\boldsymbol{\lambda}}}. \tag{A17}$$

If we then derive (A16) with respect to η and evaluate under the Gaussian null once again, we can show that

$$\frac{\mathrm{d}^3T_2(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3}=\frac{12}{\pi}\sqrt{\frac{2}{\pi}}[\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}]^{\frac{3}{2}}Z_{\tilde{\boldsymbol{\lambda}}}. \tag{A18}$$

To deal with (A4), let

$$h(\mathbf{y};\boldsymbol{\varrho})=\boldsymbol{\lambda}'\mathrm{dg}^{-1/2}(\boldsymbol{\varphi}_D)(\mathbf{y}-\boldsymbol{\varphi}_M)\eta,$$

so that $T_3(\mathbf{y};\boldsymbol{\varrho})=\log[\Phi(h)]$. A straightforward application of the chain rule implies that

$$\frac{\mathrm{d}T_3(\mathbf{y};\boldsymbol{\varrho})}{\mathrm{d}\eta}=\frac{\phi(h)}{\Phi(h)}\frac{\mathrm{d}h}{\mathrm{d}\eta}, \tag{A19}$$

$$\frac{\mathrm{d}^2T_3(\mathbf{y};\boldsymbol{\varrho})}{(\mathrm{d}\eta)^2}=\left[\frac{\phi(h)}{\Phi(h)}\right]'\left(\frac{\mathrm{d}h}{\mathrm{d}\eta}\right)^2+\frac{\phi(h)}{\Phi(h)}\frac{\mathrm{d}^2h}{(\mathrm{d}\eta)^2} \tag{A20}$$

and

$$\frac{\mathrm{d}^3T_3(\mathbf{y};\boldsymbol{\varrho})}{(\mathrm{d}\eta)^3}=\left[\frac{\phi(h)}{\Phi(h)}\right]''\left(\frac{\mathrm{d}h}{\mathrm{d}\eta}\right)^3+3\left[\frac{\phi(h)}{\Phi(h)}\right]'\frac{\mathrm{d}h}{\mathrm{d}\eta}\frac{\mathrm{d}^2h}{(\mathrm{d}\eta)^2}+\frac{\phi(h)}{\Phi(h)}\frac{\mathrm{d}^3h}{(\mathrm{d}\eta)^3}, \tag{A21}$$

where we have omitted the dependence of h on $\mathbf{y}$ and $\boldsymbol{\varrho}$ to simplify the notation.

As a consequence, we need to consider the first three derivatives of h evaluated under the Gaussian null. The first one is given by

$$\begin{aligned} \frac{\mathrm{d}h(\mathbf{y};\boldsymbol{\varrho})}{\mathrm{d}\eta} = &\ \boldsymbol{\lambda}'\frac{\mathrm{d}\ \mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\varphi}_D)}{\mathrm{d}\eta}(\mathbf{y}-\boldsymbol{\varphi}_M)\eta \\ &+\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\varphi}_D)\boldsymbol{\Omega}(\boldsymbol{\phi}_V)\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\phi}_D)\boldsymbol{\lambda}\eta+\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\varphi}_D)(\mathbf{y}-\boldsymbol{\varphi}_M), \end{aligned} \tag{A22}$$

which evaluated under the null yields

$$\frac{\mathrm{d}h(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}=\sqrt{\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}}Z_{\tilde{\boldsymbol{\lambda}}}, \tag{A23}$$

where we have used (A10) and (A15), as well as the fact that $\eta=0$.

Similarly, the derivative of (A22) with respect to η is

$$\begin{aligned}\frac{\mathrm{d}^2h(\mathbf{y};\boldsymbol{\varrho})}{(\mathrm{d}\eta)^2}=&\boldsymbol{\lambda}'\frac{\mathrm{d}^2\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{(\mathrm{d}\eta)^2}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right)\eta-\boldsymbol{\lambda}'\frac{\mathrm{d}\ \mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{\mathrm{d}\eta}\frac{\mathrm{d}\boldsymbol{\varphi}_M}{\mathrm{d}\eta}\eta+\boldsymbol{\lambda}'\frac{\mathrm{d}\ \mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{\mathrm{d}\eta}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right)\\&+\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\frac{\mathrm{d}\ \mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{\mathrm{d}\eta}\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\lambda}\eta+\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\lambda}\\&+\boldsymbol{\lambda}'\frac{\mathrm{d}\ \mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{\mathrm{d}\eta}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right)+\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)\boldsymbol{\Omega}\left(\boldsymbol{\phi}_V\right)\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\phi}_D\right)\boldsymbol{\lambda},\end{aligned} \tag{A24}$$

which, evaluated at the null simplifies to

$$\frac{\mathrm{d}^2h(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}=2\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}. \tag{A25}$$

Similarly, the derivative of (A24) with respect to η evaluated under the Gaussian null is

$$\frac{\mathrm{d}^3h(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3}=3\boldsymbol{\lambda}'\frac{\mathrm{d}^2\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{\mathrm{d}\eta^2}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right). \tag{A26}$$

Finally, we can use the fact that

$$\frac{\phi(h)}{\Phi(h)}=\sqrt{\frac{2}{\pi}},\quad\left[\frac{\phi(h)}{\Phi(h)}\right]'=-\frac{2}{\pi}\quad\text{and}\quad\left[\frac{\phi(h)}{\Phi(h)}\right]''=\frac{\sqrt{2}(4-\pi)}{\pi^{3/2}},$$

together with (A10) and (A15), to show that (A19) and (A23) imply that

$$\frac{\mathrm{d}T_3(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}=\sqrt{\frac{2}{\pi}}\sqrt{\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}}Z_{\tilde{\boldsymbol{\lambda}}}. \tag{A27}$$

In turn, (A20), (A23) and (A25) yield

$$\frac{\mathrm{d}^2T_3(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}=-\frac{2}{\pi}\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}Z^2_{\tilde{\boldsymbol{\lambda}}}+\frac{4}{\pi}\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}. \tag{A28}$$

Finally,

$$\begin{aligned}\frac{\mathrm{d}^3T_3(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3}=&\frac{4-\pi}{\pi}\sqrt{\frac{2}{\pi}}[\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}]^{\frac{3}{2}}Z^3_{\tilde{\boldsymbol{\lambda}}}-\frac{12}{\pi}\sqrt{\frac{2}{\pi}}[\boldsymbol{\lambda}'\boldsymbol{\Psi}\left(\boldsymbol{\phi}_V\right)\boldsymbol{\lambda}]^{\frac{3}{2}}Z_{\tilde{\boldsymbol{\lambda}}}\\&+3\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\frac{\mathrm{d}^2\mathrm{dg}^{-\frac{1}{2}}\left(\boldsymbol{\varphi}_D\right)}{(\mathrm{d}\eta)^2}\left(\mathbf{y}-\boldsymbol{\varphi}_M\right)\end{aligned} \tag{A29}$$

where we have used (A21), (A23), (A25) and (A26).

Given (A1), its first order condition with respect to η will be

$$\frac{\mathrm{d}l(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}=\frac{\mathrm{d}T_1(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}+\frac{\mathrm{d}T_2(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}+\frac{\mathrm{d}T_3(\mathbf{y};\boldsymbol{\varrho}_0)}{\mathrm{d}\eta}=0$$

by virtue of (A8), (A14) and (A27).

Similarly,

$$\frac{\mathrm{d}^2 l(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}=\frac{\mathrm{d}^2 T_1(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}+\frac{\mathrm{d}^2 T_2(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}+\frac{\mathrm{d}^2 T_3(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^2}=0$$

because of (A11), (A17) and (A28). Therefore, we have verified condition (A5), which guarantees Assumption 2.1 in Amengual, Bei and Sentana (2021).

As for the third derivative, if we combine (A12), (A18) and (A29), we can show that

$$\begin{aligned}\frac{\mathrm{d}^3 l(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3} &= \frac{\mathrm{d}^3 T_1(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3}+\frac{\mathrm{d}^3 T_2(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3}+\frac{\mathrm{d}^3 T_3(\mathbf{y};\boldsymbol{\varrho}_0)}{(\mathrm{d}\eta)^3}\\ &= \frac{\sqrt{2}(4-\pi)}{\pi^{\frac{3}{2}}}[\boldsymbol{\lambda}'\boldsymbol{\Psi}(\boldsymbol{\phi}_V)\boldsymbol{\lambda}]^{\frac{3}{2}}Z^3_{\tilde{\boldsymbol{\lambda}}}+3\sqrt{\frac{2}{\pi}}\boldsymbol{\lambda}'\frac{\mathrm{d}^2[\mathrm{dg}^{-\frac{1}{2}}(\boldsymbol{\varphi}_D)]}{(\mathrm{d}\eta)^2}\boldsymbol{\Psi}^{\frac{1}{2}}(\boldsymbol{\phi}_V)\mathrm{dg}^{\frac{1}{2}}(\boldsymbol{\phi}_D)\mathbf{Z}\\ &\equiv \mathcal{A}Z^3_{\tilde{\boldsymbol{\lambda}}}+\mathcal{B}'\mathbf{Z}. \end{aligned}\tag{A30}$$

Therefore, we have also verified condition (A6), so that Assumption 2.2 in Amengual, Bei and Sentana (2021) holds too.

Finally, we must purge the third derivatives in (A30) of the sampling uncertainty in estimating the mean vector and covariance matrix under the null. We can do this by orthogonalizing them with respect to the scores of the first and second moment parameters $\boldsymbol{\phi}$.

Given that

$$\frac{\partial l(\mathbf{y};\boldsymbol{\varrho}_0)}{\partial\boldsymbol{\varphi}_M}=\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)(\mathbf{y}-\boldsymbol{\varphi}_M)\equiv\mathcal{C}\mathbf{Z}$$

and

$$\frac{\partial l(\mathbf{y};\boldsymbol{\varrho}_0)}{\partial\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)}=\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V)[\mathbf{Z}'\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)\mathbf{Z}-\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)]\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varphi}_V),$$

it is easy to see that the effects of estimation uncertainty only come through $\boldsymbol{\varphi}_M$. As a consequence, we will have that the adjusted variance of $\mathrm{d}^3 l(\mathbf{y};\boldsymbol{\varrho}_0)/(\mathrm{d}\eta)^3$ will be given by

$$\begin{aligned}\mathcal{V}_{adj}=&\mathcal{V}_{\eta}-\mathcal{V}_{\eta\boldsymbol{\eta}_M}\mathcal{V}^{-1}_{\boldsymbol{\varphi}_M}\mathcal{V}_{\boldsymbol{\varphi}_M\eta}\\ =&Var\left(\mathcal{A}H_3(Z_{\tilde{\boldsymbol{\lambda}}})+\mathcal{B}'\mathbf{Z}\right)\\ &-cov\left(\mathcal{A}H_3(Z_{\tilde{\boldsymbol{\lambda}}})+\mathcal{B}'\mathbf{Z},\mathcal{C}\mathbf{Z}\right)Var^{-1}(\mathcal{C}\mathbf{Z})\,cov(\mathcal{C}\mathbf{Z},\mathcal{A}H_3(Z_{\tilde{\boldsymbol{\lambda}}})+\mathcal{B}'\mathbf{Z})\\ =&\mathcal{A}^2Var(Z^3_{\tilde{\boldsymbol{\lambda}}})\\ =&6\mathcal{A}^2.\end{aligned}$$

On this basis, Theorem 1 in Amengual, Bei and Sentana (2021) implies that

$$LR_n=\frac{1}{6}sup_{\tilde{\boldsymbol{\lambda}}\neq\mathbf{0}}\left[\sum_{i=1}^{n}H_3\left(\frac{\tilde{\boldsymbol{\lambda}}'}{\sqrt{\tilde{\boldsymbol{\lambda}}'\tilde{\boldsymbol{\lambda}}}}\mathbf{Z}_i\right)\right]^2+O_p(n^{-\frac{1}{6}}),$$

as desired. □

A.2 Proof of Proposition 2

Let $\mathbf{e}$ denote an affine transformation of the observed variables whose sample mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_K$, respectively. It is easy to see that $\mathbf{e}$ can also be written as an affine transformation of $\mathbf{y}^*=\mathbf{a}+\mathbf{B}\mathbf{y}$. Hence, expression (2) implies that the tests based on $\mathbf{y}$ and $\mathbf{y}^*$ are numerically the same. □

B Computational details of the simulations

We simulate n random draws from the multivariate skew normal distribution in (1) using the following rejection sampling method. First, we simulate $\mathbf{x} \sim N[\boldsymbol{\varphi}_M, \boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)]$ together with an independent scalar random variable u with a uniform distribution between 0 and 1. If $u \leq \Phi\left[\boldsymbol{\vartheta}'\text{dg}^{-1/2}(\boldsymbol{\varphi}_D)(\mathbf{x} - \boldsymbol{\varphi}_M)\right]$, then $\mathbf{y} = \mathbf{x}$, otherwise we discard it.

Arellano-Valle and Azzalini (2008) introduce an alternative parametrization of the multivariate skew normal distribution, which they call the central parametrization, such that the parameters of interest coincide with the means, variance and covariances of the observed variables, as well as their marginal skewness coefficients. They go from the original parametrization $(\boldsymbol{\varphi}_M, \boldsymbol{\varphi}_V, \boldsymbol{\vartheta})$ to the central one in two steps. First, they consider an intermediate vector of parameters such that

$$
\begin{aligned}
\boldsymbol{\mu} &= E(\mathbf{y}) = \boldsymbol{\varphi}_M + \boldsymbol{\tau}, \\
\boldsymbol{\Upsilon}(\boldsymbol{\upsilon}) &= V(\mathbf{y}) = \boldsymbol{\Sigma}(\boldsymbol{\varphi}_V) - \boldsymbol{\tau}\boldsymbol{\tau}', \\
\boldsymbol{\tau} &= \sqrt{\frac{2}{\pi}}\text{dg}^{1/2}(\boldsymbol{\varphi}_D)\boldsymbol{\delta}
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{\delta} &= [1 + \boldsymbol{\vartheta}'\boldsymbol{\Psi}(\boldsymbol{\varphi}_V)\boldsymbol{\vartheta}]^{-1/2}\boldsymbol{\Psi}(\boldsymbol{\varphi}_V)\boldsymbol{\vartheta}, \\
\boldsymbol{\Psi}(\boldsymbol{\varphi}_V) &= \text{dg}^{-1/2}(\boldsymbol{\varphi}_D)\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)\text{dg}^{-1/2}(\boldsymbol{\varphi}_D)
\end{aligned}
$$

and $\boldsymbol{\upsilon} = (\boldsymbol{\upsilon}_D', \boldsymbol{\upsilon}_L')'$, with $\boldsymbol{\upsilon}_D = vecd[\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})]$ and $\boldsymbol{\upsilon}_L = vecl[\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})]$. This reparametrization is a one-to-one mapping with a non-zero Jacobian determinant even at the Gaussian null. In addition, it is easy to prove that the scores corresponding to $\boldsymbol{\mu}$ evaluated at $\boldsymbol{\tau} = \mathbf{0}$ coincide with the scores corresponding to $\boldsymbol{\varphi}_M$ evaluated at $\boldsymbol{\vartheta} = \mathbf{0}$, the same being true of the scores for $\boldsymbol{\upsilon}$ and $\boldsymbol{\varphi}_V$. This is not entirely surprising in view of the fact that $\boldsymbol{\varphi}_M$ and $\boldsymbol{\Sigma}(\boldsymbol{\varphi}_V)$ directly yield $E(\mathbf{y})$ and $V(\mathbf{y})$ under normality. In contrast, all the elements of the score vector and Hessian matrix corresponding to $\boldsymbol{\tau}$ are 0 when evaluated at $\boldsymbol{\tau} = \mathbf{0}$, thereby achieving the goal of confining the singularities to those elements, as in the proof of Proposition 1. Nevertheless, the third derivatives are no longer 0. Specifically,

$$
\left.\frac{\partial^3 l}{\partial \tau_k^3}\right|_{\boldsymbol{\tau}=\mathbf{0}} = \frac{4-\pi}{2} H_{kkk}[\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Upsilon}(\boldsymbol{\upsilon})] + \frac{12}{(1-R_k^2)\upsilon_{D,k}} \left. s_{\mu_k}\right|_{\boldsymbol{\tau}=\mathbf{0}}, \tag{B31}
$$

where $H_{kkk}[\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Upsilon}(\boldsymbol{\upsilon})]$ is one of the $K(K+1)(K+2)/6$ third-order multivariate Hermite polynomials in (3) and R_k^2 is the coefficient of determination in the regression of y_k on a constant and the remaining elements of $\mathbf{y}$.

Next, Arellano-Valle and Azzalini (2008) replace each τ_k with the corresponding marginal skewness coefficient

$$
\gamma_k = \frac{E(y_k - \mu_k)^3}{\upsilon_{D,k}^{3/2}} = \frac{4-\pi}{2}\left(\frac{\tau_k}{\sqrt{\upsilon_{D,k}}}\right)^3.
$$

The problem with this reparametrization is that its first and second derivatives are 0 under the Gaussian null, but this is precisely the trick that Lee and Chesher (1986) used to re-interpret their extremum test as an LM test in the case of a single parameter. Specifically, after applying L'Hopital's rule twice, the score of γ_k evaluated at $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)' = \mathbf{0}$ is

$$
\left.\frac{\partial l}{\partial \gamma_i}\right|_{\boldsymbol{\gamma}=\mathbf{0}} = \frac{\upsilon_{D,k}}{6} H_{kkk}[\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Upsilon}(\boldsymbol{\upsilon})] + \frac{4\upsilon_{D,k}}{(4-\pi)(1-R_k^2)} \left. s_{\mu_k}\right|_{\boldsymbol{\gamma}=\mathbf{0}}, \tag{B32}
$$

which is proportional to (B31). Once we purge these derivatives from the effects of estimating the sample mean vector and covariance matrix by regressing them on the scores with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\upsilon}$ and retaining the residuals, we end up with the moment test based on $H_{kkk}[\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Upsilon}(\boldsymbol{\upsilon})]$ for $k=1,\ldots,K$.

Clearly, this procedure ignores all the other $K(K-1)(K+4)/6$ third cross-derivatives of $\boldsymbol{\tau}$ and $\boldsymbol{\gamma}$, which depend on the remaining third-order multivariate Hermite polynomials in (3).

Table 1: Monte Carlo rejection rates (in %) under null and alternative hypotheses for the multivariate Gaussian versus skew normal test

	Null hypothesis			Alternative hypotheses H_{a_1}			H_{a_2}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
Panel A: Bivariate									
$n = 400$									
GET	1.0	4.9	10.2	9.9	24.2	35.3	10.1	25.0	35.8
LM-AA	0.9	5.0	9.9	5.5	16.4	26.2	8.8	22.4	33.3
GMM	1.0	4.9	9.7	9.4	23.7	35.0	9.4	24.5	35.6
Margins	1.1	4.9	10.2	2.1	8.0	15.4	5.3	14.6	23.9
$n = 1,600$									
GET	1.0	5.3	10.4	61.9	79.4	85.5	61.8	80.4	86.8
LM-AA	1.0	5.2	10.2	30.5	54.5	65.6	47.2	70.4	79.5
GMM	1.0	5.4	9.5	56.5	77.7	85.3	56.1	77.7	85.4
Margins	1.2	5.0	9.8	5.8	16.6	25.3	22.6	43.2	55.5
Panel B: Trivariate									
$n = 400$									
GET	0.7	4.5	9.5	6.2	18.5	28.7	5.8	18.2	28.3
LM-AA	1.1	5.2	10.0	3.1	10.5	18.2	3.7	12.6	20.6
GMM	1.1	4.7	9.4	5.6	17.0	26.0	5.5	16.3	25.5
Margins	1.2	5.0	10.0	1.8	6.3	11.4	1.6	6.2	12.1
$n = 1,600$									
GET	1.0	4.9	10.0	51.6	70.7	79.9	50.6	70.6	80.2
LM-AA	1.2	5.1	9.8	12.4	28.3	39.4	18.2	37.3	48.5
GMM	0.9	4.8	9.4	38.2	61.4	71.8	37.9	61.7	72.1
Margins	1.1	5.0	9.8	2.2	8.1	14.5	3.5	10.5	18.1

Notes: Results based on 10,000 samples. Panel A and B report rejection rates for bivariate and trivariate models, respectively. The mean and variance parameters $\boldsymbol{\varphi}_M$ and $\boldsymbol{\varphi}_V$ are estimated under the null using the sample mean and covariance matrix, respectively. LM-AA denotes the Lagrange multiplier test based on the score of the skewness parameters under the parametrization proposed in Arellano-Valle and Azzalini (2008). GMM refers to the J-test based on the influence functions underlying GET. Margins denotes tests on marginal skewness –á là Jarque-Bera– for each of the components. Finite sample critical values are computed by simulation. DGPs: the true mean and covariance matrix of the generated data are set to $\mathbf{0}$ and $\mathbf{I}_k$, respectively, under both the null and alternative hypotheses. As for the alternative hypotheses, in the bivariate case $H_{a_1}: \boldsymbol{\vartheta}' = \left(\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}\right)$ and $H_{a_2}: \boldsymbol{\vartheta}' = \left(\sqrt{\frac{3}{10}}, 2\sqrt{\frac{3}{10}}\right)$, while $H_{a_1}: \boldsymbol{\vartheta}' = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ and $H_{a_2}: \boldsymbol{\vartheta}' = \left(\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)$ in the trivariate case.

Figure 1: Alignment of GET and LR under the Gaussian null

Figure 1a: Bivariate case

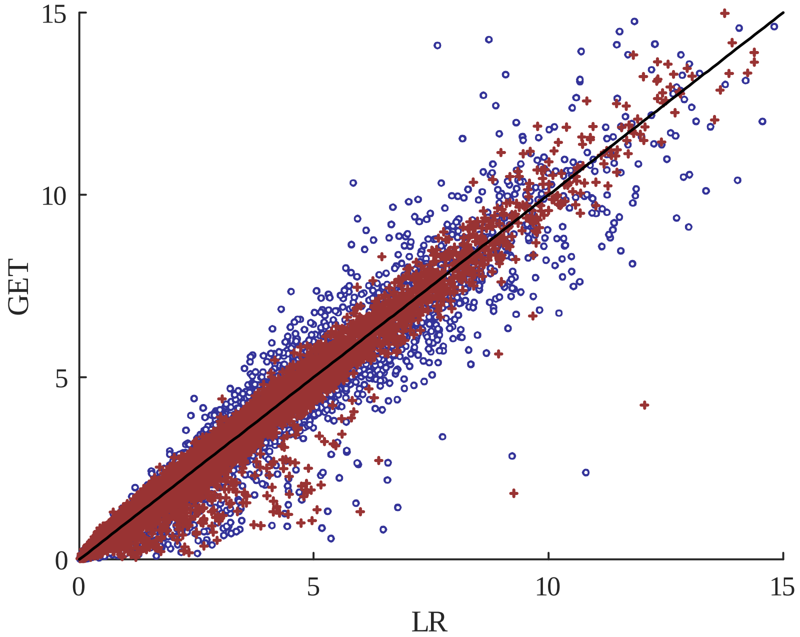
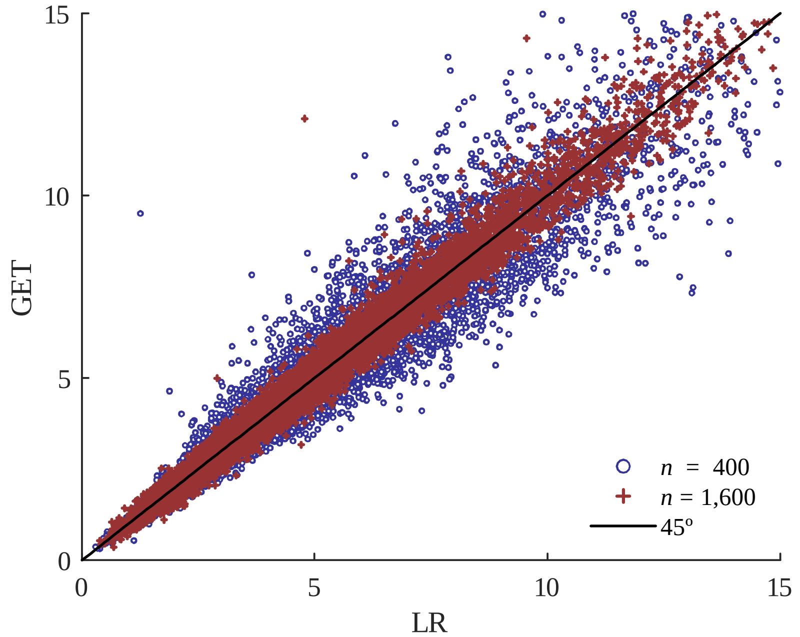


Figure 1a: Trivariate case



Notes: Scatter plots of the GET and LR test statistics based on 10,000 samples. Upper and lower panels display results for bivariate and trivariate models, respectively. The true mean and covariance matrix of the simulated Gaussian data are set to $\mathbf{0}$ and $\mathbf{I}_k$, while the mean and variance parameters $\boldsymbol{\varphi}_M$ and $\boldsymbol{\varphi}_V$ are estimated under the null using the sample mean and covariance matrix, respectively.

Figure 2: Marginal distribution of (log) city sizes for US contiguous states

Figure 2a: 2000 US census

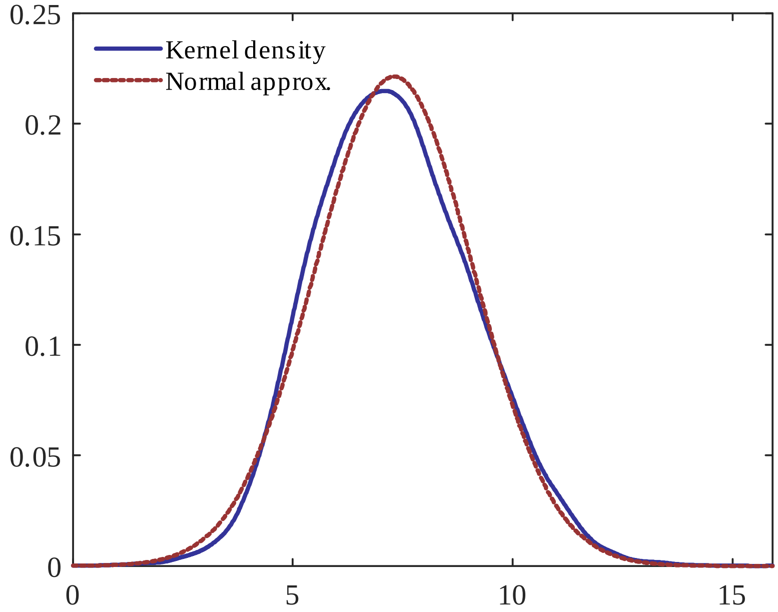
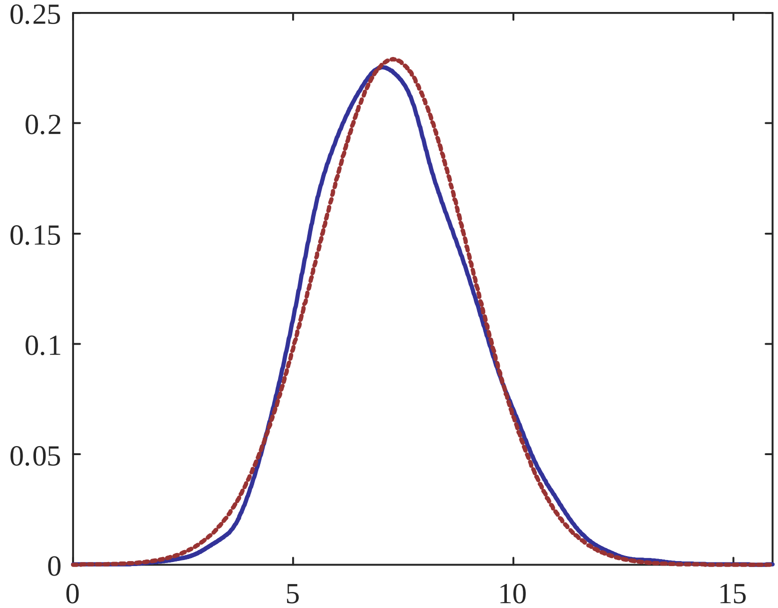
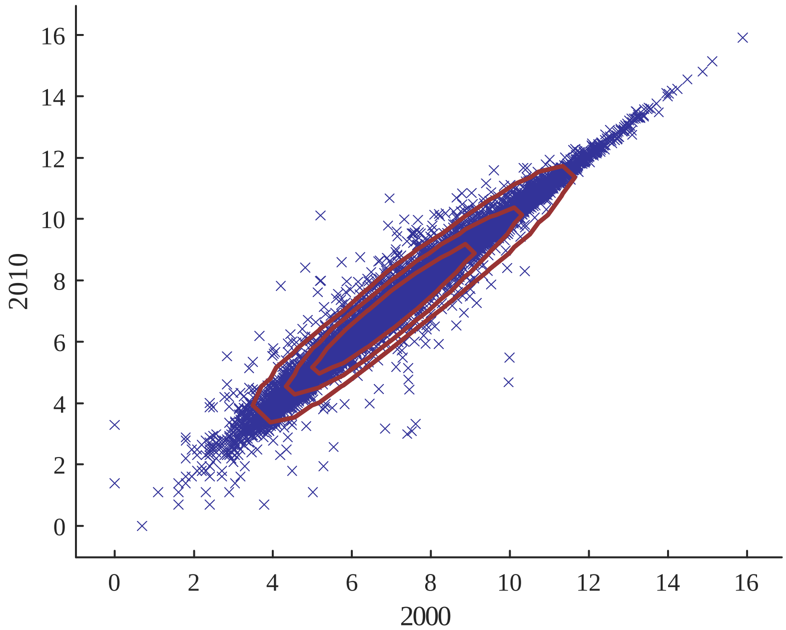


Figure 2b: 2010 US census



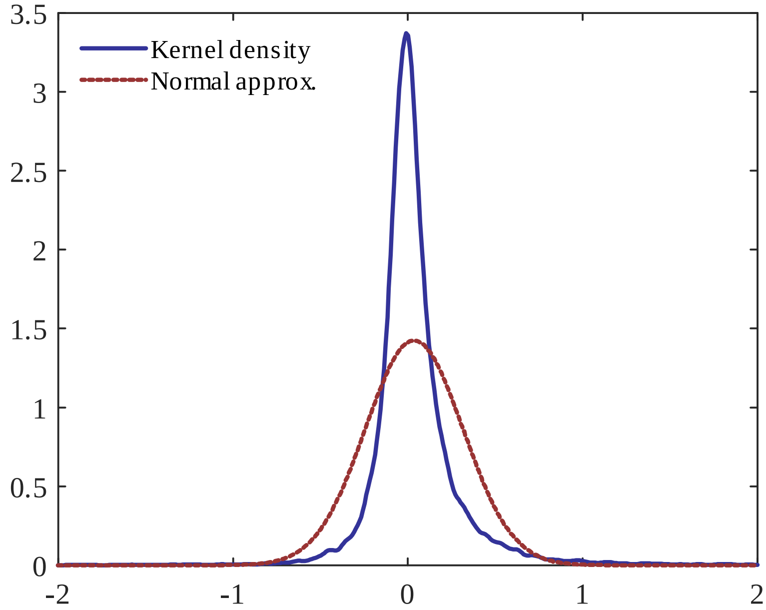
Notes: Kernel density estimates of the marginal distributions of (log) city sizes for the contiguous US states in 2000 and 2010, together with the best normal approximation to them, which share their sample means and standard deviations. We follow Eeckhout (2004) in looking at matched cities in both censuses with a population of at least one in both years and exclude Alaska, Hawaii and the remaining off-shore insular territories like Puerto Rico.

Figure 3: Bivariate distribution of (log) city sizes for the contiguous states in 2000 and 2010



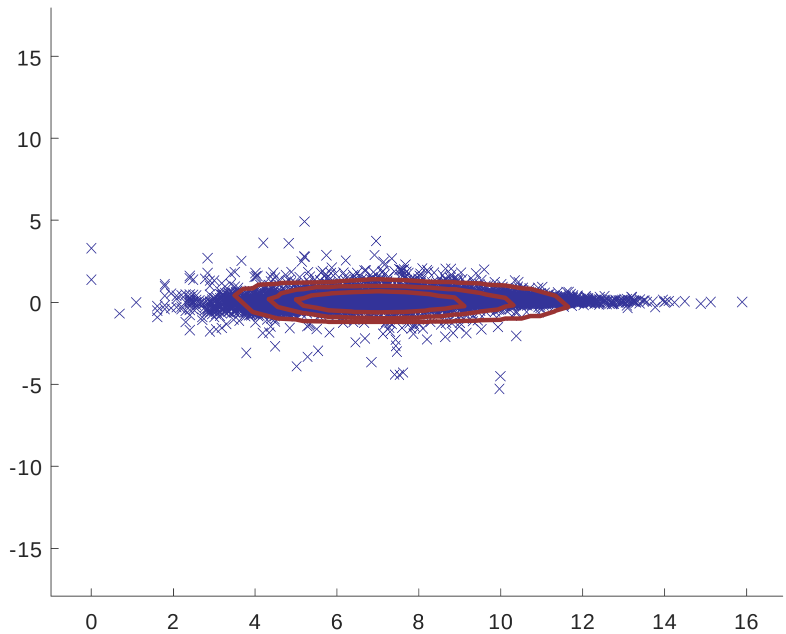
Notes: Scatter plot of (log) city sizes for the contiguous states in 2000 and 2010, together with level curves for the corresponding bivariate kernel density estimate. We follow Eeckhout (2004) in looking at matched cities in both censuses with a population of at least one in both years and exclude Alaska, Hawaii and the remaining off-shore insular territories like Puerto Rico.

Figure 4: Distribution of growth rates between 2000 and 2010



Notes: Kernel density estimates of the marginal distributions of (log) city sizes for the contiguous US states in 2000 and 2010, together with the best normal approximation to them, which share their sample means and standard deviations. We follow Eeckhout (2004) in looking at matched cities in both censuses with a population of at least one in both years and exclude Alaska, Hawaii and the remaining off-shore insular territories like Puerto Rico.

Figure 5: Joint distribution of (log) city sizes in 2000 and growth rates between 2000 and 2010.



Notes: Scatter plot of (log) city sizes for the contiguous US states in 2000 and their (geometric) growth rates between 2000 and 2010, together with level curves for the corresponding bivariate kernel density estimate. We follow Eeckhout (2004) in looking at matched cities in both censuses with a population of at least one in both years and exclude Alaska, Hawaii and the remaining off-shore insular territories like Puerto Rico.