Multivariate Hermite polynomials and information matrix tests*

Dante Amengual
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<amengual@cemfi.es>

Gabriele Fiorentini
Università di Firenze and RCEA, Viale Morgagni 59, I-50134 Firenze, Italy
<gabriele.fiorentini@unifi.it>

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

June 2021
Revised: January 2024

Abstract

The information matrix test for a normal random vector is shown to coincide with the sum of the moment tests for all third- and fourth-order multivariate Hermite polynomials. The statistic is decomposed as the sum of the marginal information matrix test for a subvector, the conditional information matrix test for the complementary subvector, and a third leftover component. It is also shown that exact finite sample distributions can be obtained by drawing spherical Gaussian vectors and orthogonalising them using sample moments. These tests are applied to assess the implications of Gibrat's law for US city sizes using the three most recent censuses.

Keywords: City size distribution, Exact tests, Hessian matrix, Likelihood factorisation, Multivariate normality, Outer product of the score.

JEL: C30, C46, C52, R12

*Corresponding author: Enrique Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. <sentana@cemfi.es>

1 Introduction

The information matrix (IM) test introduced by White (1982) constitutes a rather general procedure for examining the specification of models estimated by maximum likelihood (ML). It directly assesses the IM equality, which states that the sum of the Hessian matrix and the outer product of the score vector should be zero in expected value when the estimated model is correctly specified. As an illustration, White (1982) derived the IM test for a univariate normal random variable, proving that it simply checks that the third- and fourth-order Hermite polynomials of the standardised variable have zero means in the population. Therefore, it is equivalent to the version of the popular Jarque and Bera (1980) test proposed by Kiefer and Salmon (1983) among many others.

The theoretical properties and interpretation of the IM test as part of the general class of moment tests in Newey (1985) and Tauchen (1985) (see White (1994)), as well as its applications and finite sample behaviour, have been extensively investigated. Multivariate normality tests have also been studied extensively. The intersection is limited to Smith (1987), who related the IM test to a normality test against a multivariate Edgeworth-type A series expansion truncated to the fourth order in the context of linear simultaneous limited dependent variable models. Given both the univariate precedent in White (1982) and the results in Smith (1987), it is not surprising that we can prove that the IM test for a multivariate normal random vector coincides with the sum of the two moment tests that look at the means of all the third- and fourth-order multivariate Hermite polynomials. As a result, the IM test statistic is also equivalent to the smooth test against a fourth-order Hermite polynomial expansion of the multivariate normal density in Koziol (1987), which is in turn equivalent to Mardia and Kent's (1991) score test of multivariate normality against exponential distributions whose sufficient statistics depend not only on the levels and cross-products of the observations, but also on all possible products of three and four elements. The neglected heterogeneity interpretation of the IM test in Chesher (1984) provides a completely different justification, which might be more relevant in some empirical applications.

The numerical equivalence between the IM test and the moment test based on Hermite polynomials is important because, on the one hand, it allows the IM test, which is often regarded as a black box, to be reinterpreted in this context as a moment test of a set of rather natural influence functions. On the other hand, it provides a likelihood-based justification for using the third- and fourth-order multivariate Hermite polynomials to test normality.

Rather than in unconditional models, often the interest is in conditional models in which a subset of dependent variables is modelled as a multivariate linear regression of another subset of exogenous variables. For that reason, we deconstruct the multivariate normality test by showing that it can be computed as the sum of three asymptotically orthogonal components: a marginal IM test for the regressors, a conditional IM test for the distribution of the dependent variables given those regressors, and a third component that collects the missing terms. In turn, we show that the conditional component can be computed as the sum of the aforementioned multivariate skewness and kurtosis tests applied to the regression residuals, a multivariate regression version of White's (1980) test for conditional heteroskedasticity in those residuals, and an additional component that looks at the conditional skewness of residuals given regressors, which we call a test for conditionally heteroclicity following Bera and Lee (1993). Similarly, we also prove that the remaining component of the joint test focuses on both the conditional heteroskedasticity and heteroclicity of the regressors given the regression residuals.

We explicitly address the widespread and often justified concern that the IM test is unreliable in finite samples (see Horowitz (1994) and the reference therein) by explaining how to simulate its exact, parameter-free, finite sample distribution, as well as that of its components, to any desired degree of accuracy for any dimension of the random vector and sample size. In this respect, we exploit the numerical invariance of the different components of the IM test to affine transformations of the observed variables to simulate draws extremely quickly.

Finally, we apply our procedures to analyse the joint and conditional normality of the size of US cities and their rates of growth using data from the 2000, 2010, and 2020 censuses. As is well known, Gibrat's law says that if the (continuously compounded) rates of growth of the populations of cities are independent of their initial size, the cross-sectional distribution of city sizes in the steady state should be log-normal (see Bottazzi, Dosi, Lippi, Pammolli, and Riccaboni (2001) for a related analysis of the pharmaceutical industry).

The rest of the paper is organised as follows. Section 2 includes our results on the joint IM test. Section 3 provides the decomposition of the IM test that results from factorisation of the joint distribution into a marginal and a conditional component. The results of some Monte Carlo exercises that examine the size and power of the tests in finite samples are presented in Section 4, and that assess the joint and conditional normality of US city sizes are presented in Section 5. The conclusion in Section 6 mentions some avenues for further research. Proofs and auxiliary results are relegated to appendices.

2 The information matrix test

Our null hypothesis is that the $M \times 1$ vector is

$$\mathbf{x} \sim i.i.d.\ N(\boldsymbol{\nu}, \boldsymbol{\Gamma}) \text{ with } |\boldsymbol{\Gamma}| > 0, \tag{1}$$

and $\boldsymbol{\nu}$ and $\boldsymbol{\Gamma}$ unknown. Given a random sample on $\mathbf{x}$ of dimension N, $\{\mathbf{x}_n\}_{n=1}^N$, the maximum likelihood estimators of $\boldsymbol{\nu}$ and $\boldsymbol{\Gamma}$ coincide with the sample mean vector $\hat{\boldsymbol{\nu}}_N$ and the covariance matrix $\hat{\boldsymbol{\Gamma}}_N$ (with denominator N). If $\boldsymbol{\nu}$ and $\boldsymbol{\Gamma}$ are known, then there are no parameters to estimate under the null and, therefore, no gradient or information matrix. However, the test statistic in Proposition 1 in Section 2.3 with estimators replaced by true values would continue to be valid as a multivariate normality test. Similarly, we use the *i.i.d.* assumption mainly for computing the asymptotic variance of the influence functions, which, in principle, could be robustified for the presence of serial correlation.

2.1 Multivariate Hermite polynomials and moment tests

To enable a generalization of White's (1982) result to the multivariate context, let us follow Barndorff-Nielsen and Petersen (1979) in defining the (centred) multivariate Hermite polynomials of $\mathbf{x}$ of order $k = k_1 + \ldots + k_M \geq 0$ as

$$H_{k_1 \ldots k_M}[\boldsymbol{\varepsilon}(\boldsymbol{\nu}), \boldsymbol{\Delta}] \cdot e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\nu})'\boldsymbol{\Delta}(\mathbf{x}-\boldsymbol{\nu})} = (-1)^k \frac{\partial^k}{(\partial x_1)^{k_1} \ldots (\partial x_M)^{k_M}} \left[e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\nu})'\boldsymbol{\Delta}(\mathbf{x}-\boldsymbol{\nu})} \right], \quad (2)$$

where $\boldsymbol{\Delta} = \boldsymbol{\Gamma}^{-1}$ and $\boldsymbol{\varepsilon}(\boldsymbol{\nu}) = (\mathbf{x} - \boldsymbol{\nu})$. The mean of any Hermite polynomial of positive degree is known to be zero when model (1) is correctly specified, so it constitutes a basis for testing multivariate normality.

The symmetry of the higher-order partial derivatives in (2), however, implies that some of the M^k multivariate Hermite polynomials of order k will be replicated several times. Specifically, there are only $\binom{M+k-1}{k}$ different polynomials for a given order, so we can avoid generalised inverse matrices by eliminating the redundancies from the list of moments to test. In the third- and fourth-order cases, we can use the triplication and quadruplication matrices in Meijer (2005), which generalise the duplication matrix (see also Smith (1987) for third- and fourth-order generalisations of the duplication and elimination matrices).

For that reason, we define

$$\mathbf{H}_k(\varepsilon; \boldsymbol{\Delta}) = \begin{bmatrix} H_{k,0,\cdots,0}(\varepsilon; \boldsymbol{\Delta}) \\ H_{k-1,1,\cdots,0}(\varepsilon; \boldsymbol{\Delta}) \\ \vdots \\ H_{0,\cdots,0,k}(\varepsilon; \boldsymbol{\Delta}) \end{bmatrix} \quad (3)$$

as the $\binom{M+k-1}{k} \times 1$ vector that contains all the non-redundant multivariate Hermite polynomials of order k, which we simply denote by $\mathbf{H}_k(\boldsymbol{\varepsilon}^*)$ for the special case of $\boldsymbol{\Delta} = \mathbf{I}_M$, so that $\mathbf{H}_1(\boldsymbol{\varepsilon}^*) = \boldsymbol{\varepsilon}^*$ with $V[\mathbf{H}_1(\boldsymbol{\varepsilon}^*)] = \mathbf{I}_M$. Thus, we end up with $M(M+1)(M+2)/6$ and $M(M+1)(M+2)(M+3)/24$ distinct third- and fourth-order moment conditions, respectively, which coincide with the degrees of freedom of the asymptotic chi-square distributions under the Gaussian null of the corresponding multivariate skewness and kurtosis tests defined by

$$h_{3N} = N\,\bar{\mathbf{m}}'_{3N}(\hat{\boldsymbol{\nu}}_N, \hat{\boldsymbol{\gamma}}_N)\{V[\mathbf{H}_3(\boldsymbol{\varepsilon}^*)]\}^{-1}\bar{\mathbf{m}}_{3N}(\hat{\boldsymbol{\nu}}_N, \hat{\boldsymbol{\gamma}}_N) \tag{4}$$

and

$$h_{4N} = N\,\bar{\mathbf{m}}'_{4N}(\hat{\boldsymbol{\nu}}_N, \hat{\boldsymbol{\gamma}}_N)\{V[\mathbf{H}_4(\boldsymbol{\varepsilon}^*)]\}^{-1}\bar{\mathbf{m}}_{4N}(\hat{\boldsymbol{\nu}}_N, \hat{\boldsymbol{\gamma}}_N), \tag{5}$$

where $\boldsymbol{\gamma} = vech(\boldsymbol{\Gamma})$, $\bar{\mathbf{m}}_{3N}(\boldsymbol{\nu}, \boldsymbol{\gamma})$ and $\bar{\mathbf{m}}_{4N}(\boldsymbol{\nu}, \boldsymbol{\gamma})$ denote the sample averages of $\mathbf{H}_3[\boldsymbol{\varepsilon}^*(\boldsymbol{\nu}, \boldsymbol{\gamma})]$ and $\mathbf{H}_4[\boldsymbol{\varepsilon}^*(\boldsymbol{\nu}, \boldsymbol{\gamma})]$, respectively, over the random sample of size N, with $\boldsymbol{\varepsilon}^*(\boldsymbol{\nu}, \boldsymbol{\gamma}) = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\varepsilon}(\boldsymbol{\nu})$, and $V[\mathbf{H}_3(\boldsymbol{\varepsilon}^*)]$ and $V[\mathbf{H}_4(\boldsymbol{\varepsilon}^*)]$ denote their covariance matrices, whose theoretical expressions we provide in Lemma 2 in Section 2.3.

2.2 IM influence functions for testing multivariate normality

The contribution of one observation on $\mathbf{x}$ to the log-likelihood function is

$$-\frac{M}{2}\ln 2\pi - \frac{1}{2}\ln|\boldsymbol{\Gamma}| - \frac{1}{2}\mathbf{z}'(\boldsymbol{\nu}, \boldsymbol{\gamma})\boldsymbol{\Delta}^{-1}\mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma}),$$

where $\mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma}) = \boldsymbol{\Delta}\boldsymbol{\varepsilon}(\boldsymbol{\nu}) = \boldsymbol{\Gamma}^{-1}(\mathbf{x} - \boldsymbol{\nu})$. The scores of this component with respect to the vector of mean parameters are

$$\mathbf{s}_{\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) = \mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma}),$$

which coincide with the first-order Hermite polynomials of $\mathbf{x}$. Similarly, the scores with respect to the covariance matrix parameters are given by

$$\mathbf{s}_{\boldsymbol{\gamma}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) = \frac{1}{2}\mathbf{D}'_M vec[\mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu}, \boldsymbol{\gamma}) - \boldsymbol{\Delta}],$$

which coincide with the product of the (transposed) duplication matrix $\mathbf{D}_M$ and the second-order Hermite polynomials. Therefore, the Hessian matrix is given by

$$\begin{aligned}
\mathbf{h}_{\boldsymbol{\nu}\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) &= -\boldsymbol{\Delta}, \\
\mathbf{h}_{\boldsymbol{\gamma}\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) &= -\mathbf{D}'_M[\mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma}) \otimes \boldsymbol{\Delta}],
\end{aligned}$$

and

$$\mathbf{h}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) = -\frac{1}{2}\mathbf{D}'_M\{2[(\boldsymbol{\Delta} \otimes \mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu}, \boldsymbol{\gamma})] - (\boldsymbol{\Delta} \otimes \boldsymbol{\Delta})\}\mathbf{D}_M.$$

Hence, the sum of the outer product of the score and the Hessian, which constitute the basis for the IM test, yields the terms

$$\begin{aligned}
\mathbf{d}_{\boldsymbol{\nu}\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) &= \mathbf{s}_{\boldsymbol{\nu}\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma})\mathbf{s}'_{\boldsymbol{\nu}\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) + \mathbf{h}_{\boldsymbol{\nu}\boldsymbol{\nu}}(\mathbf{x}; \boldsymbol{\nu}, \boldsymbol{\gamma}) \qquad (6) \\
&= \mathbf{z}(\boldsymbol{\nu}, \boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu}, \boldsymbol{\gamma}) - \boldsymbol{\Delta},
\end{aligned}$$

$$
\begin{aligned}
\mathbf{d}_{\gamma\boldsymbol{\nu}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) &= \mathbf{s}_{\gamma\boldsymbol{\nu}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{s}'_{\gamma\boldsymbol{\nu}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) + \mathbf{h}_{\gamma\boldsymbol{\nu}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) \\
&= \frac{1}{2}\mathbf{D}'_M vec[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - \boldsymbol{\Delta}]\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - \mathbf{D}'_M[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma}) \otimes \boldsymbol{\Delta}],
\end{aligned} \tag{7}
$$

and

$$
\begin{aligned}
\mathbf{d}_{\gamma\gamma}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) &= \mathbf{s}_{\gamma\gamma}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{s}'_{\gamma\gamma}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) + \mathbf{h}_{\gamma\gamma}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) \\
&= \frac{1}{4}\mathbf{D}'_M vec[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - \boldsymbol{\Delta}]vec'[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma})\boldsymbol{\Delta} - \boldsymbol{\Delta}]\mathbf{D}_M \\
&\quad -\frac{1}{2}\mathbf{D}'_M\{2[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma})] - (\boldsymbol{\Delta}\otimes\boldsymbol{\Delta})\}\mathbf{D}_M.
\end{aligned} \tag{8}
$$

When model (1) is correctly specified, the IM equality holds and the mean of

$$
\mathbf{d}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) = \begin{bmatrix} \mathbf{d}_{\boldsymbol{\nu}\boldsymbol{\nu}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) \\ \mathbf{d}_{\boldsymbol{\nu}\gamma}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) \\ \mathbf{d}_{\gamma\gamma}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma}) \end{bmatrix}
$$

is zero. Hence, if we denote by $\bar{\mathbf{d}}_N(\hat{\boldsymbol{\nu}}_N,\hat{\boldsymbol{\gamma}}_N)$ the sample average of $\mathbf{d}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})$ evaluated at the ML estimators, by $V[\mathbf{d}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})]$ the covariance matrix of those influence functions adjusted for the sampling uncertainty in estimating $\boldsymbol{\nu}$ and $\boldsymbol{\gamma}$ under the null, and by $+$ the Moore-Penrose inverse of a square matrix, then the IM test of multivariate normality is simply

$$
IM_N = N\,\bar{\mathbf{d}}'_N(\hat{\boldsymbol{\nu}}_N,\hat{\boldsymbol{\gamma}}_N)\{V[\mathbf{d}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})]\}^{+}\bar{\mathbf{d}}_N(\hat{\boldsymbol{\nu}}_N,\hat{\boldsymbol{\gamma}}_N), \tag{9}
$$

which has an asymptotic chi-square distribution under the Gaussian null, with the number of degrees of freedom equal to the rank of $V[\mathbf{d}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})]$, whose singularity reflects the symmetric nature of the Hessian matrix and the corresponding outer product of the scores, the redundant nature of some of the influence functions involved, and the fact that some of them are linear functions of the scores.

2.3 Reinterpretation of the IM test

Our first result, which generalizes the example in White (1982) to the multivariate case, establishes the numerical equivalence between directly relying on (9) or using the sum of (4) and (5) for the purpose of testing the correct specification of (1).

Proposition 1 *The IM test statistic (9), which compares the outer product of the score with the Hessian of model (1) evaluated at the sample mean vector and covariance matrix, numerically coincides with the sum of the two asymptotically independent moment tests (4) and (5), which check whether the expected values of all the distinct third- and fourth-order multivariate Hermite polynomials of* $\mathbf{x}$ *are zero.*

Although we prove Proposition 1 from first principles for pedagogical reasons, it could also be derived using the results in Section 4 and Appendices A and B of Smith (1987) for the limiting case in which there are no regressors in the linear simultaneous equation limited dependent variable model that he considers, but the limited dependent variables are in fact unlimited.

Multivariate Hermite polynomials of different orders are known to be uncorrelated (see, e.g., Holmquist (1996) or Rahman (2017)), which justifies the additive decomposition of the test statistic in Proposition 1. In addition, Holly and Gardiol (1995), building on the formulas for the higher order moments of the multivariate normal in Balestra and Holly (1990), which in turn generalises Magnus and Neudecker (1979) and Phillips and Park (1988), explain how to obtain matrix expressions for the covariance matrices of the entire vector of polynomials of any given common order.

On the basis of their results, we derive computationally simple closed-form expressions for the asymptotic covariance matrices of the sample moments underlying our tests effectively adjusted for parameter uncertainty under the null of Gaussianity, which should improve the finite sample performance of our testing procedures, as forcefully argued by Orme (1990) (see also Horowitz (1994) and the references therein). Specifically, the next result contains detailed expressions for the covariances between two arbitrary first-, second-, third-, and fourth-order Hermite polynomials, thereby generalising the results in Amengual, Fiorentini, and Sentana (2022a).

Lemma 2 *Let δ_{ij} denote the $(i,j)^{th}$ element of $\mathbf{\Delta}$. When model (1) is correctly specified,*

$$
\begin{aligned}
cov(H_i, H_j) &= \delta_{ij}, \\
cov(H_{ij}, H_{i'j'}) &= \delta_{ii'}\delta_{jj'} + \delta_{ij'}\delta_{ji'}, \\
cov(H_{ijk}, H_{i'j'k'}) &= \delta_{ii'}\delta_{jj'}\delta_{kk'} + \delta_{ii'}\delta_{jk'}\delta_{kj'} + \delta_{ij'}\delta_{ji'}\delta_{kk'} \\
&\quad + \delta_{ij'}\delta_{jk'}\delta_{ki'} + \delta_{ik'}\delta_{ji'}\delta_{kj'} + \delta_{ik'}\delta_{jj'}\delta_{ki'}, \quad \textit{and} \\
cov(H_{ijkh}, H_{i'j'k'h'}) &= \delta_{ii'}\delta_{jj'}\delta_{kk'}\delta_{hh'} + \delta_{ii'}\delta_{jj'}\delta_{kh'}\delta_{hk'} + \delta_{ii'}\delta_{jk'}\delta_{kj'}\delta_{hh'} + \delta_{ii'}\delta_{jk'}\delta_{kh'}\delta_{hj'} \\
&\quad + \delta_{ii'}\delta_{jh'}\delta_{kj'}\delta_{hk'} + \delta_{ii'}\delta_{jh'}\delta_{kk'}\delta_{hj'} + \delta_{ij'}\delta_{ji'}\delta_{kk'}\delta_{hh'} + \delta_{ij'}\delta_{ji'}\delta_{kh'}\delta_{hk'} \\
&\quad + \delta_{ij'}\delta_{jk'}\delta_{ki'}\delta_{hh'} + \delta_{ij'}\delta_{jk'}\delta_{kh'}\delta_{hi'} + \delta_{ij'}\delta_{jh'}\delta_{ki'}\delta_{hk'} + \delta_{ij'}\delta_{jh'}\delta_{kk'}\delta_{hi'} \\
&\quad + \delta_{ik'}\delta_{ji'}\delta_{kj'}\delta_{hh'} + \delta_{ik'}\delta_{ji'}\delta_{kh'}\delta_{hj'} + \delta_{ik'}\delta_{jj'}\delta_{ki'}\delta_{hh'} + \delta_{ik'}\delta_{jj'}\delta_{kh'}\delta_{hi'} \\
&\quad + \delta_{ik'}\delta_{jh'}\delta_{ki'}\delta_{hj'} + \delta_{ik'}\delta_{jh'}\delta_{kj'}\delta_{hi'} + \delta_{ih'}\delta_{ji'}\delta_{kj'}\delta_{hk'} + \delta_{ih'}\delta_{ji'}\delta_{kk'}\delta_{hj'} \\
&\quad + \delta_{ih'}\delta_{jj'}\delta_{ki'}\delta_{hk'} + \delta_{ih'}\delta_{jj'}\delta_{kk'}\delta_{hi'} + \delta_{ih'}\delta_{jk'}\delta_{ki'}\delta_{hj'} + \delta_{ih'}\delta_{jk'}\delta_{kj'}\delta_{hi'}.
\end{aligned}
$$

When $\mathbf{\Gamma} = \mathbf{I}_M$, the components of $\mathbf{x}$ are stochastically independent and the multivariate Hermite polynomial $H_{k_1 \ldots k_M}[\boldsymbol{\varepsilon}(\boldsymbol{\nu}), \mathbf{\Delta}]$ simplifies to the product of the univariate polynomials $H_{k_1}[\varepsilon_1(\nu_1)]$,...,$H_{k_M}[\varepsilon_M(\nu_M)]$. Moreover, Lemma 2 implies that different multivariate Hermite polynomials of the same order become orthogonal to each other, so the IM test of model (1) effectively becomes the sum of the individual moments tests for all possible distinct multivariate Hermite polynomials of orders 3 and 4. Consequently, if we considered a sequence of local departures from a multivariate spherically normal distribution, the non-centrality parameter of the asymptotic distribution of the skewness and kurtosis tests in Proposition 1 would be the sum of the non-centrality parameters of each of the $\binom{M+2}{3} + \binom{M+3}{4}$ asymptotically independent moment tests, which is easy to compute.

In addition, the expressions for the variance terms that appear in Lemma 2 simplify considerably. Specifically, for the special case of $\mathbf{\Delta} = \mathbf{I}_M$, so that $\mathbf{H}_1(\boldsymbol{\varepsilon}^*) = \boldsymbol{\varepsilon}^*$ with $V[\mathbf{H}_1(\boldsymbol{\varepsilon}^*)] = \mathbf{I}_M$, the diagonal elements of $V[\mathbf{H}_2(\boldsymbol{\varepsilon}^*)]$ are $V(\varepsilon_i^{*2} - 1) = 2$ and $V(\varepsilon_i^* \varepsilon_{i'}^*) = 1$ for $i' \neq i$, while those of $V[\mathbf{H}_3(\boldsymbol{\varepsilon}^*)]$ are $V(\varepsilon_i^{*3} - 3\varepsilon_i^*) = 6$, $V(\varepsilon_i^{*2}\varepsilon_{i'}^* - \varepsilon_{i'}^*) = 2$ for $i' \neq i$, and $V(\varepsilon_i^* \varepsilon_{i'}^* \varepsilon_{i''}^*) = 1$ for $i'' \neq i' \neq i$. Finally, the diagonal elements of $V[\mathbf{H}_4(\boldsymbol{\varepsilon}^*)]$ are $V[(\varepsilon_i^{*2} - 3\varepsilon_i^*)^2 - 6] = 24$, $V(\varepsilon_i^{*2}\varepsilon_{i'}^{*2} - \varepsilon_i^{*2} - \varepsilon_{i'}^{*2} + 1) = 4$ for $i' \neq i$, $V(\varepsilon_i^{*3}\varepsilon_{i'}^* - 3\varepsilon_i^*\varepsilon_{i'}^*) = 6$ for $i' \neq i$, $V(\varepsilon_i^{*2}\varepsilon_{i'}^*\varepsilon_{i''}^* - \varepsilon_{i'}^*\varepsilon_{i''}^*) = 2$ for $i'' \neq i' \neq i$, and $V(\varepsilon_i^*\varepsilon_{i'}^*\varepsilon_{i''}^*\varepsilon_{i'''}^*) = 1$ for $i''' \neq i'' \neq i' \neq i$ (see Amengual, Fiorentini, and Sentana (2022a) for further details).

2.4 Computational considerations

Consider the full-rank affine transformation $\mathbf{y} = \mathbf{c} + \mathbf{D}\mathbf{x}$ with $|\mathbf{D}| \neq 0$. When (1) holds, $\mathbf{y} \sim i.i.d.\ N(\mathbf{c} + \mathbf{D}\boldsymbol{\nu}, \mathbf{D}\boldsymbol{\Gamma}\mathbf{D}')$. Our next result shows that the IM test statistic is numerically invariant to the values of $\mathbf{c}$ and $\mathbf{D}$.

Lemma 3 *The IM test statistic of model (1) numerically coincides with the analogous test statistic for* $\mathbf{y}$*.*

This numerical invariance is a very desirable property of any multivariate normality test (see Henze (2002)), but it also provides a very fast numerical procedure for computing the test statistic. Specifically, given a sample of size N on $\mathbf{x}$, we can subtract the sample mean from each observation and premultiply the resulting vector by any square root of the sample covariance matrix to create standardised random vectors for which the ML estimators of their mean vector and covariance matrix will be $\mathbf{0}$ and $\mathbf{I}_M$, respectively. Thus, the IM test statistic would be numerically equivalent to the sum of the individual moments tests for all possible multivariate Hermite polynomials of orders 3 and 4, which are very simple to compute because of their factorisation as products of univariate Hermite polynomials. Asymptotically, we can obtain the non-centrality parameter of the test for any value of $\boldsymbol{\Gamma}$ by applying the same trick.

Lemma 3 also implies that the sample mean vector and covariance matrix of the observations, which set the average of the first and second multivariate Hermite polynomials to zero, do not affect the null distribution of our proposed test in finite samples. Thus, it is possible to simulate its exact, parameter-free, finite sample distribution to any desired degree of accuracy for any dimension of $\mathbf{x}$ and sample size thanks to its pivotal nature, thereby avoiding the well-deserved criticism that the asymptotic distribution of IM tests provides a poor approximation in finite samples, especially when the number of moment conditions involved is large (see, e.g., Taylor (1987), Orme (1990), Chesher and Spady (1991), Davidson and MacKinnon (1992), and Horowitz (1994)). Specifically, it suffices to simulate R times a random sample of size N of a spherical Gaussian random vector of dimension M to obtain R independent draws of the IM test statistic for multivariate normality. Given that the sample mean and covariance matrix of a multivariate random vector take hardly any time to compute, and that the IM test statistic for random vectors standardised in the sample can also be swiftly computed, our suggested procedure generates very accurate simulated p-values very quickly. In fact, given that the only characteristics of the original sample that matter are the values of N and M, a researcher could obtain tables with exact critical values before observing the data, a very convenient strategy we follow in Sections 4 and 5.

3 Deconstructing the IM test

As we mentioned in the introduction, in empirical research the interest is often in conditional models in which a subset of dependent variables is expressed as a multivariate linear regression of another subset of exogenous variables, rather than in unconditional models. For that reason, in this section we deconstruct the multivariate normality test of Section 2 by showing that it can be computed as the sum of three asymptotically orthogonal components: a marginal IM test for the regressors, a conditional IM test for the distribution of the dependent variables given those regressors, and a third component consisting of the remaining terms, which we label as "the rest."

Specifically, the joint test we considered in the previous section assesses the correct specification of the multivariate normal distribution of $\mathbf{x}$ in (1). However, this model is known to be equivalent to

$$
\begin{aligned}
\mathbf{x}_1 &\sim i.i.d.\ N(\boldsymbol{\nu}_1, \boldsymbol{\Gamma}_1) \text{ with } |\boldsymbol{\Gamma}_1| > 0, \quad (10)\\
\mathbf{x}_2|\mathbf{x}_1 &\sim i.i.d.\ N(\boldsymbol{\alpha}_{2|1} + \mathbf{B}_{2|1}\mathbf{x}_1, \boldsymbol{\Omega}_{2|1}), \quad (11)\\
\boldsymbol{\alpha}_{2|1} &= \boldsymbol{\nu}_2 - \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\nu}_1,\\
\mathbf{B}_{2|1} &= \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}, \text{ and}\\
\boldsymbol{\Omega}_{2|1} &= \boldsymbol{\Gamma}_{22} - \boldsymbol{\Gamma}_{21}\boldsymbol{\Gamma}_{11}^{-1}\boldsymbol{\Gamma}_{21}' \text{ with } |\boldsymbol{\Omega}_{2.1}| > 0,
\end{aligned}
$$

for any conceivable partition of the M elements of $\mathbf{x}$ into two groups $\mathbf{x}_1$ and $\mathbf{x}_2$ of dimensions M_1 and M_2, respectively, with $M_1 + M_2 = M$.

Trivially, the IM test of the marginal component (10) is formally identical to the joint IM test in Proposition 1, except that it applies to $\mathbf{x}_1$ only, so all our results in Section 2 apply.

3.1 The conditional IM test: A regression interpretation

To develop the IM test of the conditional component (11), let us define $\boldsymbol{\theta}' = (\boldsymbol{\nu}_1', \boldsymbol{\gamma}_1', \boldsymbol{\theta}_{2|1}')$, $\boldsymbol{\gamma}_1 = vech(\boldsymbol{\Gamma}_1)$, $\boldsymbol{\theta}_{2|1}' = (\boldsymbol{\alpha}_{2|1}', \boldsymbol{\beta}_{2|1}', \boldsymbol{\omega}_{2|1}')$, $\boldsymbol{\beta}_{2|1} = vec(\mathbf{B}_{2|1})$, $\boldsymbol{\omega}_{2|1} = vech(\boldsymbol{\Omega}_{2|1})$,

$$\boldsymbol{\varepsilon}_{2|1}(\boldsymbol{\alpha}_{2|1}, \boldsymbol{\beta}_{2|1}) = \mathbf{x}_2 - \boldsymbol{\alpha}_{2|1} - \mathbf{B}_{2|1}\mathbf{x}_1,$$

$\boldsymbol{\Delta}_{2|1} = \boldsymbol{\Omega}_{2|1}^{-1}$, and $\boldsymbol{\varepsilon}_{2|1}^*(\boldsymbol{\theta}_{2.1}) = \boldsymbol{\Omega}_{2|1}^{-1/2}\boldsymbol{\varepsilon}_{2|1}(\boldsymbol{\alpha}_{2|1}, \boldsymbol{\beta}_{2|1})$. The derivations in Amengual, Fiorentini, and Sentana (2022b) or the results in Smith (1987) for the limiting case in which the limited dependent variables are in fact unlimited allow us to prove the following result.

Proposition 4 *The IM test that compares the outer product of the score with the Hessian of the multivariate regression model (11) evaluated at the Gaussian maximum likelihood estimators $\hat{\boldsymbol{\theta}}_N$ is asymptotically equivalent under the null hypothesis of correct specification to the sum of the four moment tests*

$$h_{hN} = N\bar{\mathbf{m}}_{hN}'(\hat{\boldsymbol{\theta}}_N)\left\{V[\mathbf{H}_2(\boldsymbol{\varepsilon}_{2|1}^*)] \otimes \begin{pmatrix} \boldsymbol{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{M_1}^+(\mathbf{I}_{M_1^2} + \mathbf{K}_{M_1M_1})(\boldsymbol{\Gamma}_1 \otimes \boldsymbol{\Gamma}_1)\mathbf{D}_{M_1}^{+\prime} \end{pmatrix}\right\}^{-1} \bar{\mathbf{m}}_{hN}(\hat{\boldsymbol{\theta}}_N), \quad (12)$$

$$h_{asN} = N\bar{\mathbf{m}}_{asN}'(\hat{\boldsymbol{\theta}}_N)\{V[\mathbf{H}_3(\boldsymbol{\varepsilon}_{2|1}^*)]\}^{-1}\bar{\mathbf{m}}_{asN}(\hat{\boldsymbol{\theta}}_N), \quad (13)$$

$$h_{acN} = N\bar{\mathbf{m}}_{acN}'(\hat{\boldsymbol{\theta}}_N)\{V[\mathbf{H}_3(\boldsymbol{\varepsilon}_{2|1}^*)] \otimes \boldsymbol{\Gamma}_1\}^{-1}\bar{\mathbf{m}}_{acN}(\hat{\boldsymbol{\theta}}_N), \quad \textit{and} \quad (14)$$

$$h_{kN} = N\bar{\mathbf{m}}_{kN}'(\hat{\boldsymbol{\theta}}_N)\{V[\mathbf{H}_4(\boldsymbol{\varepsilon}_{2|1}^*)]\}^{-1}\bar{\mathbf{m}}_{kN}(\hat{\boldsymbol{\theta}}_N), \quad (15)$$

where $\bar{\mathbf{m}}_{hN}$, $\bar{\mathbf{m}}_{asN}$, $\bar{\mathbf{m}}_{acN}$, and $\bar{\mathbf{m}}_{kN}$ are the sample averages of

$$\mathbf{m}_{hn}(\boldsymbol{\theta}) = \mathbf{H}_2[\boldsymbol{\varepsilon}_{2|1n}^*(\boldsymbol{\theta})] \otimes [(\mathbf{x}_{1n} - \boldsymbol{\nu}_1)', vech'(\mathbf{x}_{1n}\mathbf{x}_{1n}' - \boldsymbol{\Gamma}_1)]', \quad (16)$$

$$\mathbf{m}_{asn}(\boldsymbol{\theta}) = \mathbf{H}_3[\boldsymbol{\varepsilon}_{2|1n}^*(\boldsymbol{\theta})], \quad (17)$$

$$\mathbf{m}_{acn}(\boldsymbol{\theta}) = \mathbf{H}_3[\boldsymbol{\varepsilon}_{2|1n}^*(\boldsymbol{\theta})] \otimes (\mathbf{x}_{1n} - \boldsymbol{\nu}_1), \quad \textit{and} \quad (18)$$

$$\mathbf{m}_{kn}(\boldsymbol{\theta}) = \mathbf{H}_4[\boldsymbol{\varepsilon}_{2|1n}^*(\boldsymbol{\theta})], \quad (19)$$

which converge in distribution to four mutually independent chi-square random variables whose degrees of freedom are $\binom{M_2+1}{2}\frac{M_1(M_1+3)}{2}$, $\binom{M_2+2}{3}$, $\binom{M_2+2}{3}M_1$, and $\binom{M_2+3}{4}$, respectively.

Intuitively, when model (11) is correctly specified, (i) the expected value of any multivariate Hermite polynomial of positive degree k of the regression residuals conditional on the regressors is zero and (ii) the conditional covariance matrices of those polynomials coincide with the unconditional covariance mateices in Lemma 2.

In the next subsections we follow Amengual, Fiorentini, and Sentana (2022b) in providing a simple regression interpretation for each of the moment tests in Proposition 4. These interpretations in terms of Lagrange multiplier (LM) tests may prove particularly useful for the purposes of indicating the specific directions in which to focus our modelling efforts to enrich model (11).

3.1.1 Testing against conditional heteroskedasticity

Consider the multivariate regression of $\mathbf{H}_2[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]$ onto 1, $(\mathbf{x}_1 - \boldsymbol{\nu}_1)$ and $vech(\mathbf{x}_1\mathbf{x}_1' - \boldsymbol{\Gamma}_1)$. Given that (16) effectively contains the relevant normal equations of this regression evaluated under the null, it is straightforward to see that the test statistic (12) numerically coincides with the LM test of zero slopes in the aforementioned auxiliary regression (see Hall (1987) for an analogous result in the univariate case). As a consequence, if (11) holds, then the quadratic form in (12) will be asymptotically distributed as a chi-square random variable with $\binom{M_2+1}{2}\frac{M_1(M_1+3)}{2}$ degrees of freedom.

More generally, the test statistic (12) that looks at the conditional mean of the second-order multivariate Hermite polynomials can be understood as a test of neglected heterogeneity in the $\boldsymbol{\beta}_{2|1}$ parameters that determine the conditional mean of the observations, as explained by Hall (1987) and Bera and Lee (1993) in the univariate case, and Sentana (1995) in the multivariate case. Nevertheless, this test will have no power to detect time variation in the constant terms of the multivariate regression which is uncorrelated to the variation in any other of the model parameters because the first-order conditions of the estimators $\hat{\boldsymbol{\theta}}_N$ corresponding to the residual covariance matrix elements $\boldsymbol{\omega}_{2|1}$ ensure that the sample mean of $\mathbf{H}_2[\boldsymbol{\varepsilon}^*_n(\hat{\boldsymbol{\theta}}_N)]$ is zero in a regression with an intercept.

3.1.2 Testing against conditional heterocliticity and unconditional asymmetry

Consider the multivariate regression of $\mathbf{H}_3[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]$ onto a constant and $(\mathbf{x}_1 - \boldsymbol{\nu}_1)$. Given that (17) and (18) effectively provide the normal equations of this regression evaluated under the null, it is straightforward to see that (13) and (14) numerically coincide with the LM tests of zero means and zero slopes, respectively, in this auxiliary regression. In this respect, (13) converges in distribution to a chi-square random variable with $\binom{M_2+2}{3}$ degrees of freedom, while (14) will converge to an independent chi-square with $\binom{M_2+2}{3}M_1$ degrees of freedom under the Gaussian null. In fact, we can exploit this asymptotic independence to interpret the sum of (12) and (14) as a joint test of unconditional and conditional asymmetry of the regression residuals given the regressors.

If we re-write the multivariate regression model (11) in deviation from the means form as

$$\mathbf{x}_2 = \boldsymbol{\nu}_2 + \mathbf{B}_{2|1}(\mathbf{x}_1 - \boldsymbol{\nu}_1) + \boldsymbol{\Omega}_{2|1}^{1/2}\boldsymbol{\varepsilon}^*_{2|1},$$

then the results in Chesher (1984) imply that (13) is simply testing for dependence between random coefficient variation in the unconditional mean of the regressands $\boldsymbol{\nu}_2$ and the elements of the covariance matrix of the residuals $\boldsymbol{\Omega}_{2|1}$. Unlike in the previous subsection, the intercepts provide additional degrees of freedom in this case. Similarly, the test statistic (14) that examines the conditional mean of the third-order polynomials effectively assesses dependence in the neglected heterogeneity of the mean and covariance parameters $\boldsymbol{\beta}_{2|1}$ and $\boldsymbol{\omega}_{2|1}$, which in turn generate what Bera and Lee (1993) called conditional heterocliticity in the univariate case.

3.1.3 Testing against unconditional kurtosis

Consider now the multivariate regression of $\mathbf{H}_4[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]$ on a constant. Given that (19) effectively contains the normal equations of this regression evaluated under the null, it is once more straightforward to prove that the quadratic form (15) numerically coincides with the LM test of zero intercepts in this auxiliary regression. Therefore, this test statistic will be asymptotically distributed as a chi-square random variable with $\binom{M_2+3}{4}$ degrees of freedom under the null.

Using Chesher's (1984) reinterpretation of the IM test as a LM test against parameter variation once again, we can also regard the moment test statistic (15) that examines the unconditional mean of the fourth-order multivariate Hermite polynomials as a test of neglected heterogeneity in $\boldsymbol{\omega}_{2|1}$, which are the parameters that characterise the covariance matrix of the innovations, as explained by Hall (1987) in the univariate case.

Finally, it is worth mentioning that we can further exploit the asymptotic independence of the different test statistics in Proposition 4 to create a test of multivariate normality of the regression residuals $\boldsymbol{\varepsilon}^*_{2|1}$ as the sum of (13) and (15).

3.1.4 Computational considerations

From a computational point of view, it is important to emphasise that, as explained in Section 2.4, the diagonal covariance matrices of $\mathbf{H}_k(\boldsymbol{\varepsilon}^*_{2|1})$ for $k = 2, 3, 4$ do not depend on any unknown quantities under the null of correct specification. In addition, if we reconsider a full-rank affine transformation of both the dependent and independent variables given by $\mathbf{y} = \mathbf{c}+\mathbf{D}\mathbf{x}$, with $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2')'$ and $\mathbf{D}$ lower triangular of full rank, we can show the following analogue to Lemma 3.

Lemma 5 *The four components of the IM test statistic of model (11) in Proposition 4 numerically coincide with the corresponding test statistics based on* $\mathbf{y}_2$ *and* $\mathbf{y}_1$*.*

Once again, this numerical invariance provides a very fast numerical procedure for computing the test statistics in Proposition 4 because the recursive nature of the lower triangular Cholesky decomposition implies that we can systematically work with

$$\begin{bmatrix} \boldsymbol{\varepsilon}^*_{1n}(\hat{\boldsymbol{\theta}}_N) \\ \boldsymbol{\varepsilon}^*_{2|1n}(\hat{\boldsymbol{\theta}}_N) \end{bmatrix} = \begin{pmatrix} \hat{\boldsymbol{\Gamma}}_{11N} & \hat{\boldsymbol{\Gamma}}_{12N} \\ \hat{\boldsymbol{\Gamma}}_{12N}' & \hat{\boldsymbol{\Gamma}}_{22N} \end{pmatrix}^{-1/2} \begin{pmatrix} \mathbf{x}_{1n} - \hat{\boldsymbol{\nu}}_{1N} \\ \mathbf{x}_{2n} - \hat{\boldsymbol{\nu}}_{2N} \end{pmatrix} \tag{20}$$

without loss of generality. In the preceding equality, the sample mean and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively. Similarly, it is straightforward to obtain exact critical values by simulation for each of the components that appear in Proposition 4 using a procedure entirely analogous to that described in Section 2.4. In the case of the multivariate normality test of the regression residuals $\boldsymbol{\varepsilon}^*_{2|1}$ mentioned at the end of the previous subsection, our exact finite sample procedure is slightly different from the analogous procedure for testing multivariate normality of the residuals in a conditionally homoskedastic, multivariate linear regression model proposed by Dufour, Khalaf, and Beaulieu (2003) in that they treat $\mathbf{x}_2$ as fixed in repeated samples, while we also simulate $\mathbf{x}_2$. Nevertheless, they are both asymptotically valid.

Finally, the fact that the population mean and covariance matrix of $\boldsymbol{\varepsilon}^*_1$ and $\boldsymbol{\varepsilon}^*_{2|1}$ are also $\mathbf{0}$ and $\mathbf{I}_M$, respectively, implies that we can easily compute the non-centrality parameters for local deviations from the null of correct specification of model (11).

3.2 The "rest"

The sum of the IM test statistic in Proposition 1 applied to $\mathbf{x}_1$, which we call the *marginal* IM test, and the four components of the IM test statistic in Proposition 4.1, which we refer to as the *conditional* IM test, does not coincide with the IM test statistic in Proposition 1 applied to $\mathbf{x}$, which we can call the *joint* IM test. At first glance, the reason may seem to be the lack of numerical invariance of the IM to reparametrisation of the model. However, this is not the case because Amengual, Fiorentini, and Sentana (2023) show that any IM test computed using either the population version of the asymptotic covariance matrix of the influence functions or the sample version suggested by Chesher (1983) and Lancaster (1984) is numerically invariant to reparametrisation.

In fact, the real reason is that those marginal and conditional components correspond to a specific partition of the elements of $\mathbf{x}$, while the joint test considers all possible partitions.

Nevertheless, we can easily characterise the missing components.

Proposition 6 *The IM test statistic in Proposition 1 applied to* $\mathbf{x}$ *numerically coincides with the sum of the following asymptotically independent moment tests: the IM test statistic in Proposition 1 applied to the marginal model for* $\mathbf{x}_1$ *in (10), the IM statistic in Proposition 4 applied to the conditional model for* $\mathbf{x}_2$ *given* $\mathbf{x}_1$ *in (11), and the sum of the two moment tests*

$$h_{rhN} = N\,\bar{\mathbf{m}}'_{rhN}(\hat{\boldsymbol{\theta}}_N)[V[\mathbf{H}_2(\boldsymbol{\varepsilon}^*_1)]\otimes\mathbf{I}_{M_2}]^{-1}\bar{\mathbf{m}}_{rhN}(\hat{\boldsymbol{\theta}}_N) \text{ and}$$
$$h_{raN} = N\,\bar{\mathbf{m}}'_{raN}(\hat{\boldsymbol{\theta}}_N)[V[\mathbf{H}_3(\boldsymbol{\varepsilon}^*_1)]\otimes\mathbf{I}_{M_2}]^{-1}\bar{\mathbf{m}}_{raN}(\hat{\boldsymbol{\theta}}_N),$$

where $\bar{\mathbf{m}}_{rhN}$ *and* $\bar{\mathbf{m}}_{raN}$ *are the sample averages of*

$$\mathbf{m}_{rhn}(\boldsymbol{\theta}) = \mathbf{H}_2[\boldsymbol{\varepsilon}^*_{1n}(\boldsymbol{\theta})] \otimes \boldsymbol{\varepsilon}^*_{2|1n}(\boldsymbol{\theta}) \text{ and} \tag{21}$$

$$\mathbf{m}_{ran}(\boldsymbol{\theta}) = \mathbf{H}_3[\boldsymbol{\varepsilon}^*_{1n}(\boldsymbol{\theta})] \otimes \boldsymbol{\varepsilon}^*_{2|1n}(\boldsymbol{\theta}), \tag{22}$$

which converge in distribution to two mutually independent chi-square random variables whose degrees of freedom are $\binom{M_1+1}{2}M_2$ *and* $\binom{M_1+2}{3}M_2$*, respectively.*

To provide intuition for this proposition, it is convenient to exploit the numerical invariances in Lemmas 3 and 5 to focus directly on (20). The marginal component of the IM test looks at the third and fourth multivariate Hermite polynomials of $\boldsymbol{\varepsilon}^*_1$, $\mathbf{H}_3(\boldsymbol{\varepsilon}^*_1)$ and $\mathbf{H}_4(\boldsymbol{\varepsilon}^*_1)$, respectively. In turn, the conditional component focuses on the third and fourth multivariate polynomials of $\boldsymbol{\varepsilon}^*_{2|1}$, $\mathbf{H}_3(\boldsymbol{\varepsilon}^*_{2|1})$ and $\mathbf{H}_4(\boldsymbol{\varepsilon}^*_{2|1})$, the Kronecker product of its second-order polynomials $\mathbf{H}_2(\boldsymbol{\varepsilon}^*_{2|1})$ with both $\mathbf{H}_1(\boldsymbol{\varepsilon}^*_1)$ and $\mathbf{H}_2(\boldsymbol{\varepsilon}^*_1)$, as well as the Kronecker product of its third-order polynomials $\mathbf{H}_3(\boldsymbol{\varepsilon}^*_{2|1})$ with $\mathbf{H}_1(\boldsymbol{\varepsilon}^*_1)$. Therefore, the third- and fourth-order polynomials of the joint test which do not appear in either the marginal or the conditional component are $\mathbf{H}_2(\boldsymbol{\varepsilon}^*_1) \otimes \mathbf{H}_1(\boldsymbol{\varepsilon}^*_{2|1})$ and $\mathbf{H}_3(\boldsymbol{\varepsilon}^*_1) \otimes \mathbf{H}_1(\boldsymbol{\varepsilon}^*_{2|1})$, respectively, which we can interpret as focusing on the conditional heteroskedasticity and heteroclicity of $\boldsymbol{\varepsilon}^*_1$ given $\boldsymbol{\varepsilon}^*_{2|1}$. Importantly, each of the components of the conditional IM test in Proposition 4 is asymptotically independent from the marginal component in Proposition 1, as well as to the two remaining components introduced in Proposition 6, which in principle offers multiple additive aggregations.

4 Monte Carlo evidence

We conduct an extensive simulation exercise to enable an evaluation of the performance of the different tests that we discussed in previous sections. Further, we compare them with the multivariate normality tests considered by Dufour, Khalaf, and Beaulieu (2003), namely those proposed by Mardia (1970) and Kilian and Demiroglu (2000) (KD). The skewness component of Mardia's (1970) test is known to coincide with (4), while its kurtosis component is based on his proposed multivariate excess kurtosis coefficient. Given the independence of these two components in large samples in the Gaussian case, the asymptotic distribution of their sum under the null is a chi-square random variable with $M(M+1)(M+2)/6+1$ degrees of freedom. In turn, the skewness and kurtosis components of the KD test are based on the cross-sectional sum of $H_3(\varepsilon^*_i)$ and $H_4(\varepsilon^*_i)$, respectively, which means that each of them will be asymptotically distributed under normality as a chi-square random variable with M degrees of freedom. We also report their joint version, which is simply the sum of these two aggregate statistics, whose asymptotic distribution is a chi-square with $2M$ degrees of freedom.

For each design we generate 20,000 samples and consider four cross-sectional dimensions ($M = 2$, 4, 8, and 16) and three sample lengths ($N = 100$, 400, and $1,600$). To save space, we report the Monte Carlo rejection rates at the conventional 5% significance level only. We also make use of Lemmas 3 and 5 to fix the population mean vector to zero and the covariance matrix to the identity matrix, which are nevertheless freely estimated in the sample.

4.1 Size

The discussion in Sections 2.4 and 3.1.4 indicates that the finite sample size of the tests we analyse should be accurate given that we approximate the finite sample critical values with $R = 10^6$ Monte Carlo replications. Nevertheless, it is also interesting to gauge the small sample size distortions that arise when asymptotic critical values are used instead. For completeness, we also report the rejection rates obtained with simulated critical values, whose differences with the nominal values are merely due to Monte Carlo variability. In this respect, the 95% confidence interval for those rejection rates is $(4.70\%, 5.30\%)$ for 20,000 simulated samples.

The results with asymptotic critical values reported in Table 1 confirm the need for finite sample size adjustments, especially for the IM and Mardia tests when the cross-sectional dimension is large. As expected, KD is the test that shows the smallest size distortions because the number of moment conditions is linear in M, rather than cubic or quartic. When the sample length is moderately large ($N = 1,600$), the size of all tests becomes rather accurate except for the kurtosis component of the IM test. In contrast, Table 2 provides a completely different picture: Monte Carlo sizes are very accurate, with the vast majority of rejection rates within the 95% confidence set. We observe no differences across sample lengths or cross-sectional dimensions, which confirms the accuracy of the simulation-based critical values that we propose.

Finally, Table 3 reports the results on the size of the components of the IM test in Propositions 4 and 6 for the bivariate case with $N = 400$ (a sample length representative of those in our empirical application in Section 5). As explained in Section 3.1.4, we simultaneously draw $\mathbf{x}_1$ and $\mathbf{x}_2$ in each Monte Carlo simulation. The results reported in Panel A indicate that tests based on the asymptotic critical values show little size distortions, which, in any event, are corrected by the simulation-based critical values in Panel B.

4.2 Power

To assess the power properties of the several testing procedures, we generate 20,000 samples from three multivariate non-Gaussian distributions whose mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively: the asymmetric Student t distribution, the two-component location-scale mixture of normals (LSMN) discussed by Mencía and Sentana (2009), and the multivariate skew normal distribution in Azzalini and Dalla Valle (1996). Our results complement those in Best and Rayner (1988), who studied the finite sample power of Koziol's (1987) test in the bivariate case.

We again make use of Lemmas 3 and 5 to exploit skewness as a common feature for these three distributions (see Engle and Kozicki (1993)); hence, orthogonal rotations of the original random vectors in which only one variable is asymmetric can always be found. Specifically, Theorem 5.12 in Azzalini and Capitanio (2014) provides a canonical representation of the multivariate skew normal with this property. Similarly, the LSMN representation in Mencía and Sentana (2009) allows us to do the same for the other two distributions. Thus, the non-normality of the multivariate distributions is effectively governed by two parameters: the skewness and kurtosis coefficients of the only asymmetric random variable. We choose a skewness coefficient of $-\frac{3}{4}$ for all three distributions and a kurtosis coefficient of 4.5 for the two LSMNs, as the kurtosis of the skew normal is a function of its skewness parameter only (see Appendix B.3). The main difference between the skew normal distribution and the other two is that in the former, the other $M-1$ variables are Gaussian and independent, so that all the remaining third and fourth multivariate cumulants are zero, while in the latter, those variables are symmetric but neither normal nor independent of each other or of the first asymmetric component.

Table 4 reports the results corresponding to the asymmetric t distribution. As expected, power increases with the sample size N. Similarly, power increases with M except for the KD test, which does not exploit any cross third- and fourth-order moment of the non-Gaussian multivariate distribution. As we mentioned before, the IM test and the test in Mardia (1970) share the same (co-) skewness component, while the (co-) kurtosis component of the former is more powerful in all cases, except when M is small and N is simultaneously large.

The results for the LSMN distribution in Table 5 are qualitatively rather similar to those of the previous table: the KD test is the worst, while both the IM and the Mardia tests perform reasonably well. It is interesting that the IM test benefits the most from the increases in the cross-sectional dimension M.

In turn, Table 6 displays the results of the simulations with the skew normal. When the sample length is small, all tests fail to reject the null. Of more interest is that power systematically decreases with M for all sample lengths. The reason is simple. Given the canonical representation of the skew normal mentioned above, the only thing that increasing M does is to add more independent Gaussian components, which in turn add more (co-) skewness and (co-) kurtosis terms. As a result, the non-centrality parameter does not change, while the number of degrees of freedom increases. It is, therefore, not surprising that the KD test is the best performer in this case.

Finally, Table 7 displays the results of the different components of the IM test detailed in Section 3 for the case $M = 2$ and $N = 400$. Given that skewness is a common feature for the three distributions that we simulate, it is not entirely surprising that most of the power comes from the skewness component of the marginal tests. This is especially so when the distribution is skew normal, which, as expected, leads to power equal to size in all the conditional bivariate tests.

5 The distribution of US city sizes and their growth rates

We apply our procedures to analyse the joint, conditional, and marginal normality of the size of US cities and their rates of growth using the 2000, 2010, and 2020 census data. Gibrat's law says that if the (continuously compounded) rates of growth of the populations of cities are independent of their initial size, the cross-sectional distribution of city sizes in the steady state should be log-normal.

In marked contrast to earlier studies, Eeckhout (2004) forcefully argued that if one looked at the entire non-truncated sample of cities and places in the 2000 US census, their size distribution would be approximately log-normal. On the other hand, Amengual, Bei, and Sentana (2022) found that the non-normality of the joint distribution of US (log) city sizes in the 2000 and 2010 censuses was very clearly seen in their growth rates (see also Ramos (2017), and Massing, Puente-Ajovín, and Ramos (2020) for further evidence for other countries).

We extend their analysis to include the recent 2020 US census data, identifying x_2 and x_3 with the continuously compounded rates of growth between 2000 and 2010, and 2010 and 2020, respectively, and x_1 with the log city size in the 2000 census. Thus, we can simultaneously study not only the joint distribution of initial city sizes and their rates of growth, whose independence is at the core of Gibrat's law, but also the relationship between two consecutive growth rates.

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. 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 of the Census Designated Places (CDPs) in the 2000, 2010, and 2020 censuses files. For that reason, we look at the joint distribution of the matched cities with a population of at least one in each of the censuses, as in Eeckhout (2004). Some CDPs were redefined or merged during our sample period, which results in anomalously high rises or drops in the population figures. Moreover, the values reported by the US Census Bureau are incorrect for a handful of CDPs, but we could not find reliable figures from other sources. For these reasons, we removed 32 outlier observations from the merged sample, so that the effective sample size contains 23,830 observations. Consequently, the average number of observations across states is equal to 496. The median value is 383 and the interquartile range is 381, with a minimum of 22 CDPs in Rhode Island and maximum of 1,443 in Texas.

Figure 1 displays scatter plots for the three different pairs that we can form with x_1, x_2, and x_3 for the 48 contiguous states. We also include kernel density estimates of the marginal distributions for these three variables, together with the best normal approximations to them, which share their sample means and standard deviations. As highlighted by Eeckhout (2004), the estimated density of (log) city sizes for the contiguous states in 2000 does not differ much from its normal approximation. Specifically, there is little evidence of kurtosis and only some evidence of asymmetry around the mode of the distribution rather than at the tails. The marginal normality test for this univariate distribution confirms both these impressions, with a kurtosis coefficient of 3.02, which is not statistically significantly different from 3, and a skewness coefficient of 0.27, which is nevertheless statistically significant in view of the large number of observations.

In contrast, the joint bivariate distributions look rather non-normal. In fact, the joint IM tests for each pair, as well as for the three variables together, reject massively, with p-values on the order of 10^{-5}. Part of the reason is probably that the variables are significantly positively correlated with each other (0.15, 0.26, and 0.24 in Figures 1a, 1b, and 1c, respectively), which contradicts the main assumption underlying Gibrat's law (see also Ishikawa et al. (2020)).

Given that in the last few decades interstate migrations in the US have become less frequent than in the past, we also conducted the analysis at the state level. Table 3 reports the number of states that reject the various components of the IM statistics that we discussed in Section 3 at the 5% level, with exact critical values computed for each test statistic using one million simulated samples for the appropriate number of cities. Specifically, by sequentially conditioning x_2 on x_1, and x_3 on x_1 and x_2, we can look at the following:

1. normality of (log) city sizes in 2000 (Panel A), which in turn, we decompose into its skewness and kurtosis components;

2. normality of the rate of growth between 2010 and 2000 conditional on (log) city sizes in 2000 (Panel B), which we also decompose into the different components highlighted in Proposition 4;

3. the residual of the joint normality test for x_1 and x_2 in Proposition 6 (Panel C);

4. joint normality of the rate of growth between 2010 and 2000 conditional on (log) city sizes in 2000 (Panel D);

5. normality of the rate of growth between 2020 and 2010 conditional on the rate of growth between 2010 and 2000 and (log) city sizes in 2000 (Panel E), which once again we decompose along the lines of Proposition 4;
6. the residual of the joint normality test for the three variables in Proposition 6 (Panel F); and
7. joint normality of the three variables (Panel G).

Panel A confirms that (log) city sizes within states differ from normality mainly through asymmetry, with weaker evidence of kurtosis. In contrast, when we analyse the conditional distribution of the rate of growth between 2010 and 2000 given (log) city size in 2000 in Panel B using the conditionally homoskedastic, linear regression model of x_2 on a constant and x_1, the different null hypotheses are rejected in almost all states, except for conditional homoskedasticity and conditional symmetry, against which we find little evidence in a few states. Interestingly, the leftover component in Proposition 6 reported in Panel C does not reject for more than half the states, so the joint normality results in Panel D are mainly driven by those in Panel B.

In turn, the pattern of rejections for the conditional distribution of the rate of growth between 2020 and 2010 given both the rate of growth between 2010 and 2000 and (log) city sizes in 2000 in Panel E is qualitatively similar to that in Panel B, indicating the presence of non-normality, conditional heteroskedasticity and conditional heteroclicity in the residuals of the conditionally homoskedastic, linear regression model of x_3 on a constant, x_1 and x_2. Moreover, the leftover term of Proposition 6 reported in Panel F leads to conclusions similar to those in Panel C. Finally, the conclusions for the bivariate and trivariate normality tests in Panel D and Panel G, respectively, also agree, which is not entirely surprising given that they reflect the sum of all the other components.

6 Conclusions

We have shown that the IM test for a normal random vector coincides with the sum of the moment tests for all third- and fourth-order multivariate Hermite polynomials. We have also decomposed this joint test as the sum of the marginal IM test for a subvector, the conditional IM test for the complementary subvector, and a third leftover component. In turn, the conditional IM test is the sum of an analogous multivariate normality test for the regression residuals, the multivariate version of White's test for conditional homoskedasticity, and a test for conditionally homoclicity which assesses the potential dependence of the third-order multivariate Hermite polynomials of those residuals on the regressors. Finally, we decompose the leftover component as the sum of analogous tests for conditional homoskedasticity and conditional homoclicity of the regressors given the regression residuals.

We also show that all these tests are numerically invariant to affine transformations of the variables involved, which considerably simplifies their calculation and also implies that they are pivotal in finite samples. As a result, we can simulate exact finite sample distributions in no time by drawing many spherical Gaussian vectors and orthogonalising them using sample moments.

Finally, we use all these tests to assess the implications of Gibrat's law for US city sizes using the three most recent censuses, finding that although the marginal distribution of (log) city sizes is reasonably close to a normal, their (continuously compounded) growth rates are not independent of either past growth rates or initial city sizes.

Our Monte Carlo exercises confirm the non-trivial power of the IM tests against empirically plausible alternatives, even though they are not consistent, because in arbitrary large samples they would fail to reject with probability one departures from normality such that all third- and fourth-order cumulants are zero. Unlike in the univariate case, the construction mechanism for distributions with this characteristic is not obvious because it is difficult to ensure the global positivity of multivariate Hermite expansions of the Gaussian density.

The IM test can be extended to examine the correct specification of more general multivariate distributions. Amengual, Fiorentini, and Sentana (2023) are currently exploring this interesting research avenue for finite Gaussian mixtures.

Acknowledgements

We would like to thank Xinyue Bei, Alberto Holly and Richard Smith for helpful comments and discussions, and Grant Hillier for his useful suggestions regarding the proofs of Proposition 1 and Lemma 2, as well as Appendix B.1. We are also grateful to Jan Eeckhout and Rafael González-Val for sharing their data with us, and to Tao Wang for able research assistance. An associate editor and two anonymous referees provided valuable feedback too, but, of course, the usual caveat applies. The first and third authors gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00, while the second one is thankful to MIUR through the PRIN project "High-dimensional time series for structural macroeconomic analysis in times of pandemic".

References

Amengual, D., Bei, X. and Sentana, E. (2022): "Normal but skewed?", *Journal of Applied Econometrics* 37, 1295–1313.

Amengual, D., Fiorentini, G. and Sentana, E. (2022a): "Moment tests of independent components", *SERIEs* 13, 429–474.

Amengual, D., Fiorentini, G. and Sentana, E. (2022b): "Tests for random coefficient variation in vector autoregressive models", in J.J. Dolado, L. Gambetti and C. Matthes (eds.) *Essays in honour of Fabio Canova*, *Advances in Econometrics* 44B, 1-35, Emerald.

Amengual, D., Fiorentini, G. and Sentana, E. (2023): "Information matrix tests for Gaussian mixtures", mimeo, CEMFI.

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.

Balestra, P. and Holly, A. (1990): "A general Kronecker formula for the moments of the multivariate normal distribution", DEEP Cahier 9002, University of Lausanne.

Barndorff-Nielsen, O. and Petersen, B.V. (1979): "The bivariate Hermite polynomials up to order six", *Scandinavian Journal of Statistics* 6, 127–128.

Bera, A. and Lee, S (1993): "Information matrix test, parameter heterogeneity and ARCH: a synthesis", *Review of Economic Studies* 60, 229–240.

Best, D.J. and Rayner, J.C.W. (1988): "A test for bivariate normality", *Statistics and Probability Letters* 6, 407–412.

Bottazzi, G., Dosi, G., Lippi, M., Pammolli, F. and Riccabon, M. (2001): "Innovation and corporate growth in the evolution of the drug industry", *International Journal of Industrial Organization* 19, 1161–1876.

Chesher, A. (1983): "The information matrix test: simplified calculation via a score test interpretation", *Economics Letters* 13, 45–48.

Chesher, A. (1984): "Testing for neglected heterogeneity", *Econometrica* 52, 865–872.

Chesher, A. and Spady, R. (1991): "Asymptotic expansions of the information matrix test statistic", *Econometrica* 59, 787–815.

Davidson, R. and MacKinnon, J.G. (1992): "A new form of the information matrix test", *Econometrica* 60, 145–157.

Dufour, J.-M., Khalaf, L. and Beaulieu, M.-C. (2003) "Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models", *Oxford Bulletin of Economics and Statistics* 65, 891–906.

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.

Fiorentini, G. and Sentana, E. (2021): "Specification tests for non-Gaussian maximum likelihood estimators", *Quantitative Economics* 12, 683-742.

Hall, A. (1987): "The information matrix test for the linear model", *Review of Economic Studies* 54, 257–263.

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

Holly, A. and Gardiol, L. (1995): An asymptotic expansion for the distribution of test criteria which are asymptotically distributed as chi-squared under contiguous alternatives", chap 5 in G. S. Maddala, P.C.B. Phillips and T. N. Srinivasan (eds) *Advances in Econometrics and Quantitative Economics: Essays in honor of Professor C. R. Rao*, Blackwells.

Holmquist, B. (1996): "The d-variate vector Hermite polynomial of order k", *Linear Algebra and its Applications* 237/238, 155–190.

Horowitz, J. (1994): "Bootstrap-based critical values for the information matrix test", *Journal of Econometrics* 61, 395–411.

Ishikawa, A., Fujimoto, S., Ramos, A. and Mizuno, T. (2020): "Initial value dependence of urban population's growth-rate distribution and the long-term growth", *Frontiers in Physics* 8, 1–10.

Jarque, C.M. and Bera, A.K. (1980): "Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals", *Economic Letters* 6, 255–259.

Kiefer, N.M. and Salmon, M. (1983): "Testing normality in econometric models", *Economic Letters* 11, 123–127.

Kilian, L., and Demiroglu, U. (2000): "Residual-based test for normality in autoregressions: asymptotic theory and simulation evidence", *Journal of Business and Economic Statistics* 18, 40–50.

Koziol, J.A. (1987): "An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives", *Metrika* 34, 17–24.

Lancaster, A. (1984): "The covariance matrix of the information matrix test", *Econometrica* 52, 1051–1053.

Magnus, J.R. (1988): *Linear structures*, Oxford University Press.

Magnus, J.R. and Neudecker, H. (1979): "The commutation matrix: some properties and applications", *Annals of Statistics* 7, 381–394.

Magnus, J.R. and Neudecker, H. (2019): *Matrix differential calculus with applications in Statistics and Econometrics*, 3rd edition, Wiley.

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

metrika 57, 519–530.

Mardia, K. V. and Kent, J.T. (1991): "Rao score tests for goodness of fit and independence", *Biometrika* 78, 355–363.

Massing, T., Puente-Ajovín, M. and Ramos, A. (2020): "On the parametric description of log-growth rates of cities' sizes of four European countries and the USA", *Physica A* 551, 1–11.

Meijer, E. (2005): "Matrix algebra for higher order moments", *Linear Algebra and its Applications* 410, 112–134.

Mencía, J. and Sentana, E. (2009): "Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation", *Journal of Econometrics* 153, 105–121.

Newey, W.K. (1985): "Maximum likelihood specification testing and conditional moment tests", *Econometrica* 53, 1047–70.

Orme, C. (1990): "The small-sample performance of the information-matrix test", *Journal of Econometrics* 46, 309–331.

Phillips, P. C. B. and Park, J.Y. (1988): "On the formulation of Wald tests of nonlinear restrictions", *Econometrica* 56, 1065–1083.

Rahman, S. (2017): "Wiener–Hermite polynomial expansion for multivariate Gaussian probability measures", *Journal of Mathematical Analysis and Applications* 454, 303–334.

Ramos, A. (2017): "Are the log-growth rates of city sizes distributed normally? Empirical evidence for the USA", *Empirical Economics* 53, 1109–1123.

Sentana, E. (1995): "Quadratic ARCH models", *Review of Economic Studies* 62, 639-661.

Smith, R.J. (1987): "Testing the normality assumption in multivariate simultaneous limited dependent variable models", *Journal of Econometrics* 34, 105–123.

Tauchen, G. (1985): "Diagnostic testing and evaluation of maximum likelihood models", *Journal of Econometrics* 30, 415–443.

Taylor, L.W. (1987): "The size bias of White's information matrix test", *Economics Letters* 24, 63–67.

White, H. (1980): "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity", *Econometrica* 48, 817–838.

White, H. (1982): "Maximum likelihood estimation of misspecified models", *Econometrica* 50, 1–25.

White, H. (1994): *Estimation, inference and specification analysis*, Cambridge University Press.

A Proofs

A.1 Proof of Proposition 1

If we vectorise the expressions (6)–(8) before we premultiply or postmultiply them by the duplication matrix or its transpose and ignore the dependence of $\mathbf{z}(\boldsymbol{\nu}, \gamma)$ on $\boldsymbol{\nu}$ and $\boldsymbol{\gamma}$ for notational simplicity, then we obtain that the $\boldsymbol{\nu\nu}$ block of the sum of the outer product of the score with the Hessian will be

$$vec(\mathbf{zz}' - \boldsymbol{\Delta}) = (\mathbf{z} \otimes \mathbf{z}) - \boldsymbol{\delta}, \tag{A1}$$

where $\boldsymbol{\delta} = vec(\boldsymbol{\Delta})$, because

$$vec(\mathbf{zz}') = (\mathbf{z} \otimes \mathbf{z}).$$

Similarly, the $\boldsymbol{\gamma\nu}$ block will be related to

$$vec[vec(\mathbf{zz}' - \boldsymbol{\Delta})\mathbf{z}' - 2(\mathbf{z} \otimes \boldsymbol{\Delta})] = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}) - (\mathbf{z} \otimes \boldsymbol{\delta}) - 2(\mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{z} \otimes \boldsymbol{\delta}), \tag{A2}$$

where $\mathbf{K}_{MM}$ is the commutation matrix of orders M and M, because

$$\begin{aligned} vec[vec(\mathbf{zz}')\mathbf{z}'] &= [\mathbf{z} \otimes vec(\mathbf{zz}')] = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec[vec(\boldsymbol{\Delta})\mathbf{z}'] &= (\mathbf{z} \otimes \boldsymbol{\delta}) \text{ and} \\ vec(\mathbf{z} \otimes \boldsymbol{\Delta}) &= (1 \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{z} \otimes \boldsymbol{\delta}) = (\mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{z} \otimes \boldsymbol{\delta}), \end{aligned}$$

in view of Theorem 3.10 in Magnus and Neudecker (2019).

Finally, the $\boldsymbol{\gamma\gamma}$ block will depend on

$$\begin{aligned} & vec\{vec(\mathbf{zz}' - \boldsymbol{\Delta})vec'(\mathbf{zz}' - \boldsymbol{\Delta}) - [4(\boldsymbol{\Delta} \otimes \mathbf{zz}') - 2(\boldsymbol{\Delta} \otimes \boldsymbol{\Delta})]\} \\ = & (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}) - (\mathbf{z} \otimes \mathbf{z} \otimes \boldsymbol{\delta}) - 5(\boldsymbol{\delta} \otimes \mathbf{z} \otimes \mathbf{z}) + (\boldsymbol{\delta} \otimes \boldsymbol{\delta}) + 2(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\boldsymbol{\delta} \otimes \boldsymbol{\delta}) \end{aligned} \tag{A3}$$

because

$$\begin{aligned} vec[vec(\mathbf{zz}')vec'(\mathbf{zz}')] &= [vec(\mathbf{zz}') \otimes vec(\mathbf{zz}')] = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec[\boldsymbol{\delta} vec'(\mathbf{zz}')] &= [vec(\mathbf{zz}') \otimes \boldsymbol{\delta}] = (\mathbf{z} \otimes \mathbf{z} \otimes \boldsymbol{\delta}), \\ vec[vec(\mathbf{zz}')\boldsymbol{\delta}'] &= [\boldsymbol{\delta} \otimes vec(\mathbf{zz}')] = (\boldsymbol{\delta} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec(\boldsymbol{\delta\delta}') &= (\boldsymbol{\delta} \otimes \boldsymbol{\delta}), \\ vec(\boldsymbol{\Delta} \otimes \mathbf{zz}') &= (\mathbf{I}_M \otimes \mathbf{K}_{1M} \otimes \mathbf{I}_M)[\boldsymbol{\delta} \otimes vec(\mathbf{zz}')] = (\boldsymbol{\delta} \otimes \mathbf{z} \otimes \mathbf{z}) \text{ and} \\ vec(\boldsymbol{\Delta} \otimes \boldsymbol{\Delta}) &= (\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\boldsymbol{\delta} \otimes \boldsymbol{\delta}). \end{aligned}$$

Holly and Gardiol (1995) express the vectors of first, second, third and fourth centred multivariate Hermite polynomials of $\mathbf{z}$ in matrix notation as

$$\mathbf{S}_{M\boldsymbol{\iota}_1}\mathbf{z}$$

$$\mathbf{S}_{M\boldsymbol{\iota}_2}[(\mathbf{z}\otimes\mathbf{z})-\boldsymbol{\delta}], \tag{A4}$$

$$\mathbf{S}_{M\boldsymbol{\iota}_3}[(\mathbf{z}\otimes\mathbf{z}\otimes\mathbf{z})-3(\mathbf{z}\otimes\boldsymbol{\delta})] \text{ and} \tag{A5}$$

$$\mathbf{S}_{M\boldsymbol{\iota}_4}[(\mathbf{z}\otimes\mathbf{z}\otimes\mathbf{z}\otimes\mathbf{z})-6(\mathbf{z}\otimes\mathbf{z}\otimes\boldsymbol{\delta})+3(\boldsymbol{\delta}\otimes\boldsymbol{\delta})], \tag{A6}$$

where $\mathbf{S}_{M\boldsymbol{\iota}_k}$ $(k=1,\ldots,4)$ are the symmetrisation operators discussed by Holmquist (1996), whose detailed expressions we provide in Supplemental Appendix B.1. In this respect, the vectors $\mathbf{H}_k$ in (3) for $k=1,\ldots,4$ contain the non-redundant elements of these expressions.

As the detailed analysis of the bivariate case in Appendix B.2.2 illustrates, the sum of the outer product of the scores and the Hessian matrix contains either duplicated elements or others which are multiples of each other. Premultiplying or postmultiplying by the (transpose of the) duplication matrix eliminates some of those duplicities, but not all of them. For that reason, in the rest of the proof we will show that the symmetrised values of (A1), (A2) and (A3) are 0 in expectation by showing that they coincide with (A4), (A5) and (A6), respectively.

It is easy to see that the $\boldsymbol{\nu\nu}$ term coincides with the second-order Hermite polynomials because $\mathbf{S}_{M\boldsymbol{\iota}_2}$ applied to $(\mathbf{z}\otimes\mathbf{z})$ has no effect and $\mathbf{K}_{MM}\boldsymbol{\delta}=\boldsymbol{\delta}$ by the symmetry of $\boldsymbol{\Delta}$. However, a comparison of this term with $\mathbf{s}_{\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})$ confirms that these cannot be used for testing purposes because they will be identically 0 when evaluated at the ML estimators when the mean and variance parameters are freely estimated.

Let us now look at the $\boldsymbol{\gamma\nu}$ block. Clearly, $\mathbf{S}_{M\boldsymbol{\iota}_3}$ applied to $(\mathbf{z}\otimes\mathbf{z}\otimes\mathbf{z})$ has no effect either. In contrast, if we apply $6\mathbf{S}_{M\boldsymbol{\iota}_3}$ to $(\mathbf{z}\otimes\boldsymbol{\delta})$ we obtain

$$\begin{aligned}
&[\mathbf{I}_{M^3}+(\mathbf{I}_M\otimes\mathbf{K}_{MM})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)+(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)\\
&+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)](\mathbf{z}\otimes\boldsymbol{\delta})\\
&=(\mathbf{z}\otimes\boldsymbol{\delta})+(\mathbf{z}\otimes\boldsymbol{\delta})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})\\
&+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})\\
&=2[(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})],
\end{aligned}$$

so that

$$\begin{aligned}
(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{z}\otimes\boldsymbol{\delta}) &= (\mathbf{z}\otimes\boldsymbol{\delta}) \text{ and}\\
(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta}) &= \mathbf{K}_{M^2M}(\mathbf{z}\otimes\boldsymbol{\delta})=(\boldsymbol{\delta}\otimes\mathbf{z})
\end{aligned}$$

by virtue of Theorems 3.7 (iii) and 3.1 in Magnus (1988), and

$$(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)](\mathbf{z}\otimes\boldsymbol{\delta})=(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\boldsymbol{\delta}\otimes\mathbf{z})=(\boldsymbol{\delta}\otimes\mathbf{z}).$$

Similarly,

$$
\begin{aligned}
6\mathbf{S}_{M\boldsymbol{\iota}_3}(\boldsymbol{\delta}\otimes\mathbf{z}) = [&\mathbf{I}_{M^3}+(\mathbf{I}_M\otimes\mathbf{K}_{MM})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)+(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_N)\\
&+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)](\boldsymbol{\delta}\otimes\mathbf{z})\\
&=(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\boldsymbol{\delta}\otimes\mathbf{z})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\boldsymbol{\delta}\otimes\mathbf{z})\\
&+(\mathbf{z}\otimes\boldsymbol{\delta})+(\mathbf{z}\otimes\boldsymbol{\delta})\\
&=2[(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\boldsymbol{\delta}\otimes\mathbf{z})],
\end{aligned}
$$

because

$$
\begin{aligned}
(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\boldsymbol{\delta}\otimes\mathbf{z}) &= (\boldsymbol{\delta}\otimes\mathbf{z}) \text{ and}\\
(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\boldsymbol{\delta}\otimes\mathbf{z}) &= \mathbf{K}_{MM^2}(\boldsymbol{\delta}\otimes\mathbf{z})=(\mathbf{z}\otimes\boldsymbol{\delta})
\end{aligned}
$$

by virtue of expression (3.3) in Magnus (1988), which implies that $\mathbf{K}_{MM^2}=\mathbf{K}_{M^2M}^{-1}$, and his Theorem 3.1.

Finally,

$$
\begin{aligned}
6\mathbf{S}_{M\boldsymbol{\iota}_3}(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta}) = [&\mathbf{I}_{M^3}+(\mathbf{I}_M\otimes\mathbf{K}_{MM})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)+(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)\\
&+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{I}_M\otimes\mathbf{K}_{MM})(\mathbf{K}_{MM}\otimes\mathbf{I}_M)](\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})\\
&=(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{z}\otimes\boldsymbol{\delta})+(\mathbf{z}\otimes\boldsymbol{\delta})\\
&+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})\\
&=2[(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})].
\end{aligned}
$$

because

$$
(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{K}_{MM}\otimes\mathbf{I}_M)=\mathbf{I}_{M^3}.
$$

Hence,

$$
\begin{aligned}
&\mathbf{S}_{M\boldsymbol{\iota}_3}[(\mathbf{z}\otimes\boldsymbol{\delta})+2(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})]\\
=\;&[(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{MM}\otimes\mathbf{I}_M)(\mathbf{z}\otimes\boldsymbol{\delta})]=3\mathbf{S}_{M\boldsymbol{\iota}_3}(\mathbf{z}\otimes\boldsymbol{\delta}),
\end{aligned}
$$

so that $\mathbf{S}_{M\boldsymbol{\iota}_3}$ times (A2) does indeed coincide with (A5). In effect, the proof is exploiting expression (B16) in Appendix B.1 below.

An entirely analogous procedure confirms that if one premultiplies (A3) by $\mathbf{S}_{M\boldsymbol{\iota}_4}$, one ends up with (A6) by virtue of expression (B17) and the fact that

$$
\mathbf{S}_{M\boldsymbol{\iota}_4}(\boldsymbol{\delta}\otimes\mathbf{z}\otimes\mathbf{z})=\mathbf{S}_{M\boldsymbol{\iota}_4}(\mathbf{z}\otimes\mathbf{z}\otimes\boldsymbol{\delta})
$$

because both the left- and right-hand side expressions involve all possible permutations of the same vectors. □

A.2 Proof of Lemma 2

It follows directly from Proposition 8 in Rahman (2017). □

A.3 Proof of Lemma 3

Given that the mapping from $\mathbf{x}$ to $\mathbf{y}$ is affine, its first-order Jacobian will be $\mathbf{B}$ while all other higher-order Jacobians will be 0. As a result, the application of Faà di Bruno's generalised chain rule to (2) implies that the vector of multivariate Hermite polynomials of order k for $\mathbf{y}$ will be $\mathbf{B}^{\otimes k} = \underbrace{\mathbf{B} \otimes \mathbf{B} \otimes ... \otimes \mathbf{B}}_{k \text{ times}}$ times the vector of multivariate Hermite polynomials of order k for $\mathbf{x}$. The numerical invariance of moment tests to linear transformations of the influence functions with constant coefficients yields the desired result. □

A.4 Proof of Proposition 4

Given (11), the conditional mean vector and covariance matrix of $\mathbf{x}_2$ given $\mathbf{x}_1$ will be

$$\boldsymbol{\mu}_2(\boldsymbol{\theta}_{2|1}) = \boldsymbol{\alpha}_{2|1} + \mathbf{B}_{2|1}\mathbf{x}_1 = \mathbf{\Pi}_{2|1}\mathbf{w}_1 \quad \text{and} \quad \mathbf{\Sigma}_2(\boldsymbol{\theta}) = \mathbf{\Omega}_{2|1},$$

respectively, where $\mathbf{w}_1' = (1, \mathbf{x}_1')$, $\mathbf{\Pi}_{2|1} = (\boldsymbol{\alpha}_{2|1}|\mathbf{B}_{2|1})$ and $\boldsymbol{\pi}_{2|1} = vec(\mathbf{\Pi}_{2|1})$, so that $\boldsymbol{\theta}_{2|1} = (\boldsymbol{\pi}_{2|1}', \boldsymbol{\omega}_{2|1}')$. For simplicity of notation, we shall drop the 2|1 subscripts in what follows. Consequently, the contribution from a single observation n to the conditional log-likelihood function is

$$-\frac{M_2}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}_{2.1}| - \frac{1}{2}(\mathbf{x}_{2n} - \mathbf{\Pi}\mathbf{w}_{1n})'\mathbf{\Omega}^{-1}(\mathbf{x}_{2n} - \mathbf{\Pi}\mathbf{w}_{1n}) = -\frac{M}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{\Omega}| - \frac{1}{2}\varsigma_n(\boldsymbol{\theta}),$$

where $\varsigma_n(\boldsymbol{\theta}) = \boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})$.

The maximum likelihood estimators of the model parameters are known in closed-form without the need to conduct any numerical optimisation. Specifically,

$$(\hat{\boldsymbol{\alpha}}_N, \hat{\mathbf{B}}_N) = \hat{\mathbf{\Pi}}_N = \left(\sum_{n=1}^{N} \mathbf{x}_{2n}\mathbf{w}_{1n}'\right)\left(\sum_{n=1}^{N} \mathbf{w}_{1n}\mathbf{w}_{1n}'\right)^{-1}$$

and

$$\hat{\mathbf{\Omega}}_N = \frac{1}{N}\left[\sum_{n=1}^{N}(\mathbf{x}_{2n} - \hat{\mathbf{\Pi}}_N\mathbf{w}_{1n})(\mathbf{x}_{2n} - \hat{\mathbf{\Pi}}_N\mathbf{w}_{1n})'\right].$$

Nevertheless, we need expressions for the score and Hessian matrix to be able to derive the IM test.

To compute the score, we first differentiate $\boldsymbol{\mu}_n(\boldsymbol{\theta})$ and $\mathbf{\Sigma}_n(\boldsymbol{\theta})$ with respect to the $q = M_2(M_1 + 1) + M_2(M_2+1)/2$ model parameters in $\boldsymbol{\theta}$. Specifically, the first derivatives are given by

$$\begin{aligned} \frac{\partial \boldsymbol{\mu}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\pi}'} &= \mathbf{w}_{1n}' \otimes \mathbf{I}_{M_2} \text{ and} \\ \frac{\partial vec[\mathbf{\Sigma}_n(\boldsymbol{\theta})]}{\partial \boldsymbol{\omega}'} &= \mathbf{D}_{M_2}. \end{aligned}$$

Thus, the conditional log-likelihood score is

$$\mathbf{s}_n(\boldsymbol{\theta}) = \mathbf{w}_{ln}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta}) + \mathbf{Z}_{sn}(\boldsymbol{\theta})vec[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_{M_2}],$$

where

$$\begin{aligned}
\mathbf{Z}_{ln}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{w}_{1n} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime} \\ \mathbf{0} \end{bmatrix} \text{ and} \\
\mathbf{Z}_{sn}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{0} \\ \frac{1}{2}\mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-\frac{1}{2}\prime} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime}) \end{bmatrix}.
\end{aligned}$$

As a result, the scores will be

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\pi}n}(\boldsymbol{\theta}) &= [\mathbf{w}_{1n} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime}\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})] = \left[\begin{pmatrix} 1 \\ \mathbf{x}_{1n} \end{pmatrix} \otimes \boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})\right] \\
&= vec[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})\mathbf{x}'_{1n}] \qquad \text{(A7)}
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\omega}n}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-\frac{1}{2}\prime} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime})vec[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_M] \\
&= \frac{1}{2}\mathbf{D}'_{M_2}vec[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}]. \qquad \text{(A8)}
\end{aligned}$$

Consequently, the outer product of the scores will be

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\pi}n}(\boldsymbol{\theta})\mathbf{s}'_{\boldsymbol{\pi}n}(\boldsymbol{\theta}) &= [\mathbf{w}_{1n}\mathbf{w}'_{1n} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime}\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-\frac{1}{2}}] \\
&= [\mathbf{w}_{1n}\mathbf{w}'_{1n} \otimes \boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}],
\end{aligned}$$

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\omega}n}(\boldsymbol{\theta})\mathbf{s}'_{\boldsymbol{\pi}n}(\boldsymbol{\theta}) &= \frac{1}{2}\mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-\frac{1}{2}\prime} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime})vec[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_M][\mathbf{w}'_{1n} \otimes \boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Omega}^{-\frac{1}{2}}] \\
&= \frac{1}{2}\mathbf{D}'_{M_2}vec[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}][\mathbf{w}'_{1n} \otimes (\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}]
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{s}_{\boldsymbol{\omega}n}(\boldsymbol{\theta})\mathbf{s}'_{\boldsymbol{\omega}n}(\boldsymbol{\theta}) &= \frac{1}{4}\mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-\frac{1}{2}\prime} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime})vec[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_{M_2}] \\
&\quad \times vec'[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta}) - \mathbf{I}_{M_2}](\boldsymbol{\Omega}^{-\frac{1}{2}\prime} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}\prime})\mathbf{D}_{M_2} \\
&= \frac{1}{4}\mathbf{D}'_{M_2}vec[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}] \\
&\quad \times vec'[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n} - \boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}]\mathbf{D}_{M_2}.
\end{aligned}$$

To compute the Hessian, it is convenient to use the general expressions for elliptical distributions in Supplementary Appendix C of Fiorentini and Sentana (2021), namely

$$\mathbf{h}_{\boldsymbol{\theta\theta}n}(\boldsymbol{\phi}) = \frac{\partial^2 d_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 g\left[\varsigma_n(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{(\partial\varsigma)^2}\frac{\partial\varsigma_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\varsigma_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} + \frac{\partial g\left[\varsigma_n(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial\varsigma}\frac{\partial^2\varsigma_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},$$

where

$$\partial^2 d_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\mathbf{Z}_{sn}(\boldsymbol{\theta})\mathbf{Z}'_{sn}(\boldsymbol{\theta}) - \frac{1}{2}\left\{vec'\left[\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\theta})\right]\otimes\mathbf{I}_q\right\}\partial vec\left\{\partial vec'\left[\boldsymbol{\Sigma}_n(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}\right\}/\partial\boldsymbol{\theta}'$$

and

$$\begin{aligned}\partial^2\varsigma_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' \quad = \quad & 2\mathbf{Z}_{ln}(\boldsymbol{\theta})\mathbf{Z}'_{ln}(\boldsymbol{\theta}) + 8\mathbf{Z}_{sn}(\boldsymbol{\theta})[\mathbf{I}_{M_2}\otimes\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})]\mathbf{Z}'_{sn}(\boldsymbol{\theta})\\ & +4\mathbf{Z}_{ln}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_{M_2}]\mathbf{Z}'_{sn}(\boldsymbol{\theta}) + 4\mathbf{Z}_{sn}(\boldsymbol{\theta})[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\otimes\mathbf{I}_{M_2}]\mathbf{Z}'_{ln}(\boldsymbol{\theta})\\ & -2[\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_n^{-\frac{1}{2}\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_q]\partial vec[\partial\boldsymbol{\mu}'_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}]\partial\boldsymbol{\theta}'\\ & -\{vec'[\boldsymbol{\Sigma}_n^{-\frac{1}{2}}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\boldsymbol{\Sigma}_n^{-\frac{1}{2}\prime}(\boldsymbol{\theta})]\otimes\mathbf{I}_q\}\partial vec\{\partial vec'[\boldsymbol{\Sigma}_n(\boldsymbol{\theta})]/\partial\boldsymbol{\theta}\}/\partial\boldsymbol{\theta}'.\end{aligned}$$

In the case of model (11), $d_n(\boldsymbol{\theta}) = -\frac{1}{2}\ln|\boldsymbol{\Omega}|$ and

$$\partial^2 d_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = \frac{1}{2}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1})\mathbf{D}_{M_2}\end{bmatrix}.$$

Similarly, we have that $g\left[\varsigma_n(\boldsymbol{\theta}),\boldsymbol{\eta}\right] = -\frac{1}{2}\varsigma_n(\boldsymbol{\theta})$ under normality, so that $\partial g\left[\varsigma_n(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/\partial\varsigma = -\frac{1}{2}$ and $\partial^2 g\left[\varsigma_n(\boldsymbol{\theta}),\boldsymbol{\eta}\right]/(\partial\varsigma)^2 = 0$. Finally,

$$\begin{aligned}\partial^2\varsigma_n(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = 2\begin{pmatrix}\mathbf{x}_{1n}\mathbf{x}'_{1n}\otimes\boldsymbol{\Omega}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\\ +2\left\{\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-\frac{1}{2}\prime}\otimes\boldsymbol{\Omega}^{-\frac{1}{2}\prime})[\mathbf{I}_{M_2}\otimes\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})](\boldsymbol{\Omega}^{-\frac{1}{2}}\otimes\boldsymbol{\Omega}^{-\frac{1}{2}})\mathbf{D}_{M_2}\end{array}\right\}\\ +2\left\{\begin{array}{cc}\mathbf{0} & (\mathbf{x}_{1n}\otimes\boldsymbol{\Omega}^{-\frac{1}{2}\prime})[\boldsymbol{\varepsilon}_n^{*\prime}(\boldsymbol{\theta})\otimes\mathbf{I}_{M_2}](\boldsymbol{\Omega}^{-\frac{1}{2}}\otimes\boldsymbol{\Omega}^{-\frac{1}{2}})\mathbf{D}_{M_2}\\ \mathbf{0} & \mathbf{0}\end{array}\right\}\\ +2\left\{\begin{array}{cc}\mathbf{0} & \mathbf{0}\\ \mathbf{D}'_{M_2}(\boldsymbol{\Omega}^{-\frac{1}{2}\prime}\otimes\boldsymbol{\Omega}^{-\frac{1}{2}\prime})[\boldsymbol{\varepsilon}_n^*(\boldsymbol{\theta})\otimes\mathbf{I}_{M_2}](\mathbf{x}'_{1n}\otimes\boldsymbol{\Omega}^{-\frac{1}{2}}) & \mathbf{0}\end{array}\right\}\\ = 2\left\{\begin{array}{cc}(\mathbf{w}_{1n}\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1}) & [\mathbf{w}'_{1n}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}]\mathbf{D}_{M_2}\\ \mathbf{D}'_{M_2}[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1}] & \mathbf{D}'_{M_2}[\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}]\mathbf{D}_{M_2}\end{array}\right\},\end{aligned}$$

where we have exploited the fact that the second derivatives of the conditional mean and covariance functions with respect to the model parameters are all zero.

Therefore, we can write the Hessian matrix as

$$-\left\{\begin{array}{c}(\mathbf{w}_{1n}\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1})\\ \mathbf{D}'_{M_2}[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1}]\end{array}\right.\left.\begin{array}{c}[\mathbf{w}'_{1n}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}]\mathbf{D}_{M_2}\\ \mathbf{D}'_{M_2}\{\boldsymbol{\Omega}^{-1}\otimes[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}-\frac{1}{2}\boldsymbol{\Omega}^{-1}]\}\mathbf{D}_{M_2}\end{array}\right\}$$

The sum of the outer product of the score and the Hessian yields the following three terms:

$$\boldsymbol{\pi\pi}\text{: } [\mathbf{w}_{1n}\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}] - (\mathbf{w}_{1n}\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1}), \qquad \text{(A9)}$$

$$\begin{aligned}\boldsymbol{\omega\pi} \quad : \quad & \frac{1}{2}\mathbf{D}'_{M_2}vec[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}][\mathbf{w}'_{1n}\otimes(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}]\\ & -\mathbf{D}'_{M_2}[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})\mathbf{w}'_{1n}\otimes\boldsymbol{\Omega}^{-1}], \qquad \text{(A10)}\end{aligned}$$

and

$$\boldsymbol{\omega\omega}: \frac{1}{4}\mathbf{D}'_{M_2} vec[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}]$$
$$\times\, vec'[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}-\boldsymbol{\Omega}^{-1}]\mathbf{D}_{M_2}$$
$$-\mathbf{D}'_{M_2}\{\boldsymbol{\Omega}^{-1}\otimes[\boldsymbol{\Omega}^{-1}(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})(\mathbf{x}_{2n}-\boldsymbol{\Pi}\mathbf{w}_{1n})'\boldsymbol{\Omega}^{-1}-\frac{1}{2}\boldsymbol{\Omega}^{-1}]\}\mathbf{D}_{M_2}. \quad \text{(A11)}$$

When $\mathbf{x}_{1n} = 1$, these formulas reduce to those in the proof of Proposition 1. In fact, a straightforward application of the arguments in that proof eventually show that the expressions for the symmetrised version of the sum of the Hessian and the outer product of the scores coincide with the influence functions $\mathbf{m}_{hn}(\boldsymbol{\theta})$, $\mathbf{m}_{asn}(\boldsymbol{\theta})$, $\mathbf{m}_{acn}(\boldsymbol{\theta})$ and $\mathbf{m}_{kn}(\boldsymbol{\theta})$. Therefore, the only task left is to derive expressions for the asymptotic covariance matrices of the sample averages of those influence functions. But since we are maintaining the assumption of *i.i.d.* sampling, and the conditional distribution of the standardised regression residuals does not depend on the regressors under the null, we can easily prove that

$$\lim_{N\to\infty} V[\sqrt{N}\bar{\mathbf{m}}_{hN}(\hat{\boldsymbol{\theta}}_N)] = V\{\mathbf{H}_2[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]\}\otimes\begin{bmatrix} 0 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}^+_{M_1}(\mathbf{I}_{M_1^2}+\mathbf{K}_{M_1M_1})(\boldsymbol{\Gamma}_1\otimes\boldsymbol{\Gamma}_1)\mathbf{D}^{+\prime}_{M_1}\end{bmatrix}, \quad \text{(A12)}$$

$$\lim_{N\to\infty} V[\sqrt{N}\bar{\mathbf{m}}_{asN}(\hat{\boldsymbol{\theta}}_N)] = V\{\mathbf{H}_3[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]\}, \quad \text{(A13)}$$

$$\lim_{N\to\infty} V[\sqrt{N}\bar{\mathbf{m}}_{acN}(\hat{\boldsymbol{\theta}}_N)] = V\{\mathbf{H}_3[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]\}\otimes\boldsymbol{\Gamma}_1 \text{ and} \quad \text{(A14)}$$

$$\lim_{N\to\infty} V[\sqrt{N}\bar{\mathbf{m}}_{kN}(\hat{\boldsymbol{\theta}}_N)] = V\{\mathbf{H}_4[\boldsymbol{\varepsilon}^*_{2|1}(\boldsymbol{\theta})]\}, \quad \text{(A15)}$$

where the only slight complication is to prove that

$$V\{[1,(\mathbf{x}_1-\boldsymbol{\nu}_1)', vech'(\mathbf{x}_1\mathbf{x}_1'-\boldsymbol{\Gamma}_1)]'\} = \begin{bmatrix} 0 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Gamma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}^+_{M_1}(\mathbf{I}_{M_1^2}+\mathbf{K}_{M_1M_1})(\boldsymbol{\Gamma}_1\otimes\boldsymbol{\Gamma}_1)\mathbf{D}^{+\prime}_{M_1}\end{bmatrix},$$

which follows directly from the expressions for the third- and fourth-order central moments of a multivariate normal random vector with zero mean and covariance matrix $\boldsymbol{\Gamma}_1$. □

A.5 Proof of Lemma 5

The proof follows immediately from well-known numerical invariance properties of multivariate regression residuals to lower triangular affine transformations of the regressors and the regressands. □

A.6 Proof of Proposition 6

Given the numerical invariance of the test statistics in Lemmas 2 and 5, the proof of this statement can be obtained by comparing the influence functions involved in Propositions 1 and 4 after transforming the observations using the population version of (20). □

B Auxiliary results and computational details

B.1 The symmetrisation operators

The correct expressions for the first four symmetrisation operators discussed by Holmquist (1996) are

$$\mathbf{S}_{M\boldsymbol{\iota}_1} = \mathbf{I}_M,$$
$$\mathbf{S}_{M\boldsymbol{\iota}_2} = \frac{1}{2}(\mathbf{I}_{M^2} + \mathbf{K}_{MM}),$$
$$\mathbf{S}_{M\boldsymbol{\iota}_3} = \frac{1}{6}[\mathbf{I}_{M^3} + (\mathbf{I}_M \otimes \mathbf{K}_{MM}) + (\mathbf{K}_{MM} \otimes \mathbf{I}_M) + (\mathbf{I}_M \otimes \mathbf{K}_{MM})(\mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+(\mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_M \otimes \mathbf{K}_{MM}) + (\mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_M \otimes \mathbf{K}_{MM})(\mathbf{K}_{MM} \otimes \mathbf{I}_M)] \text{ and}$$
$$\mathbf{S}_{M\boldsymbol{\iota}_4} = \frac{1}{24}[\mathbf{I}_{M^4} + (\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM}) + (\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M) + (\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM}) + (\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2}) + (\mathbf{K}_{MM} \otimes \mathbf{K}_{MM}) + (\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})$$
$$+(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2}) + (\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{K}_{MM})$$
$$+(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})$$
$$+(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M) + (\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+\mathbf{K}_{M^2M^2} + (\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})\mathbf{K}_{M^2M^2} + (\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})$$
$$+(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)$$
$$+(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{K}_{MM})$$
$$+(\mathbf{I}_{M^2} \otimes \mathbf{K}_{MM})(\mathbf{K}_{MM} \otimes \mathbf{I}_{M^2})(\mathbf{I}_M \otimes \mathbf{K}_{MM} \otimes \mathbf{I}_M)(\mathbf{K}_{MM} \otimes \mathbf{K}_{MM})$$
$$+(\mathbf{K}_{MN} \otimes \mathbf{I}_{N^2})\mathbf{K}_{N^2N^2} + (\mathbf{I}_{N^2} \otimes \mathbf{K}_{NN})(\mathbf{K}_{NN} \otimes \mathbf{I}_{N^2})\mathbf{K}_{N^2N^2}.$$

The adjectival noun "symmetrisation" reflects the fact that when one applies these operators to the arbitrary vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ of dimension M, one ends up with

$$\mathbf{S}_{M\boldsymbol{\iota}_1}\mathbf{a} = \mathbf{a},$$
$$\mathbf{S}_{M\boldsymbol{\iota}_2}(\mathbf{a}\otimes\mathbf{b}) = \frac{1}{2}[(\mathbf{a}\otimes\mathbf{b}) + (\mathbf{b}\otimes\mathbf{a})],$$
$$\mathbf{S}_{M\boldsymbol{\iota}_3}(\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}) = \frac{1}{6}[(\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}) + (\mathbf{a}\otimes\mathbf{c}\otimes\mathbf{b}) + (\mathbf{b}\otimes\mathbf{a}\otimes\mathbf{c})$$
$$+(\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{a}) + (\mathbf{c}\otimes\mathbf{a}\otimes\mathbf{b}) + (\mathbf{c}\otimes\mathbf{b}\otimes\mathbf{a})],$$
$$\mathbf{S}_{M\boldsymbol{\iota}_4}(\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{d}) = \frac{1}{24}[(\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{d}) + (\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{d}\otimes\mathbf{c}) + (\mathbf{a}\otimes\mathbf{c}\otimes\mathbf{b}\otimes\mathbf{d}) + (\mathbf{a}\otimes\mathbf{c}\otimes\mathbf{d}\otimes\mathbf{b})$$
$$+(\mathbf{a}\otimes\mathbf{d}\otimes\mathbf{b}\otimes\mathbf{c}) + (\mathbf{a}\otimes\mathbf{d}\otimes\mathbf{c}\otimes\mathbf{b}) + (\mathbf{b}\otimes\mathbf{a}\otimes\mathbf{c}\otimes\mathbf{d}) + (\mathbf{b}\otimes\mathbf{a}\otimes\mathbf{d}\otimes\mathbf{c})$$
$$+(\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{a}\otimes\mathbf{d}) + (\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{d}\otimes\mathbf{a}) + (\mathbf{b}\otimes\mathbf{d}\otimes\mathbf{a}\otimes\mathbf{c}) + (\mathbf{b}\otimes\mathbf{d}\otimes\mathbf{c}\otimes\mathbf{a})$$
$$+(\mathbf{c}\otimes\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{d}) + (\mathbf{c}\otimes\mathbf{a}\otimes\mathbf{d}\otimes\mathbf{b}) + (\mathbf{c}\otimes\mathbf{b}\otimes\mathbf{a}\otimes\mathbf{d}) + (\mathbf{c}\otimes\mathbf{b}\otimes\mathbf{d}\otimes\mathbf{a})$$
$$+(\mathbf{c}\otimes\mathbf{d}\otimes\mathbf{a}\otimes\mathbf{b}) + (\mathbf{c}\otimes\mathbf{d}\otimes\mathbf{b}\otimes\mathbf{a}) + (\mathbf{d}\otimes\mathbf{a}\otimes\mathbf{b}\otimes\mathbf{c}) + (\mathbf{d}\otimes\mathbf{a}\otimes\mathbf{c}\otimes\mathbf{b})$$
$$+(\mathbf{d}\otimes\mathbf{b}\otimes\mathbf{a}\otimes\mathbf{c}) + (\mathbf{d}\otimes\mathbf{b}\otimes\mathbf{c}\otimes\mathbf{a}) + (\mathbf{d}\otimes\mathbf{c}\otimes\mathbf{a}\otimes\mathbf{b}) + (\mathbf{d}\otimes\mathbf{c}\otimes\mathbf{b}\otimes\mathbf{a})].$$

Two very useful properties of these operators that Grant Hillier has shared with us are

$$\mathbf{S}_{M\boldsymbol{\iota}_3}(\mathbf{K}_{MM}\otimes\mathbf{I}_M) = \mathbf{S}_{M\boldsymbol{\iota}_3} \text{ and} \tag{B16}$$
$$\mathbf{S}_{M\boldsymbol{\iota}_4}(\mathbf{I}_M\otimes\mathbf{K}_{MM}\otimes\mathbf{I}_M) = \mathbf{S}_{M\boldsymbol{\iota}_4}, \tag{B17}$$

which effectively follow from the fact that postmultiplying by $(\mathbf{K}_{MM}\otimes\mathbf{I}_M)$ and $(\mathbf{I}_M\otimes\mathbf{K}_{MM}\otimes\mathbf{I}_M)$ just rearranges the terms in $\mathbf{S}_{M\boldsymbol{\iota}_3}$ and $\mathbf{S}_{M\boldsymbol{\iota}_4}$, respectively.

B.2 Special cases

B.2.1 The univariate case

The contribution of x to the log-likelihood function is

$$-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\gamma^2 - \frac{\varepsilon^2(\nu)}{2\gamma^2}$$

The score of this component with respect to the mean parameter is

$$s_\nu(x;\nu,\gamma^2) = z(\nu,\gamma^2),$$

while the score with respect to the variance parameter is given by

$$s_{\boldsymbol{\gamma}^2}(x;\nu,\gamma) = \frac{1}{2}[z^2(\nu,\gamma^2) - \delta^2],$$

where $\delta^2 = \gamma^{-2}$, so they coincide with the first and second Hermite polynomials of $z(\nu,\gamma^2)$.

In turn, the Hessian matrix is given by

$$\begin{bmatrix} h_{\boldsymbol{\nu\nu}}(x;\nu,\gamma^2) & h_{\boldsymbol{\gamma\nu}}(x;\nu,\gamma^2) \\ h_{\boldsymbol{\gamma\nu}}(x;\nu,\gamma^2) & h_{\boldsymbol{\gamma\gamma}}(x;\nu,\gamma) \end{bmatrix} = -\begin{bmatrix} \delta^2 & \delta^2 z(\nu,\gamma^2) \\ \delta^2 z(\nu,\gamma^2) & \delta^2[z^2(\nu,\gamma^2) - \delta^2] \end{bmatrix},$$

while the covariance matrix of the score will be the expected value of the outer product matrix

$$\left[\begin{array}{cc} z^2(\nu,\gamma^2) & \frac{1}{2}z(\nu,\gamma^2)[z^2(\nu,\gamma^2)-\delta^2] \\ \frac{1}{2}z(\nu,\gamma^2)[z^2(\nu,\gamma^2)-\delta^2] & \frac{1}{4}[z^2(\nu,\gamma^2)-\delta^2]^2 \end{array}\right].$$

Therefore, the sum of the outer product of the score and the Hessian yields the following three terms

$$\begin{aligned} \nu\nu &: \; z^2(\nu,\gamma^2)-\delta^2, \\ \gamma^2\nu &: \; \frac{1}{2}z(\nu,\gamma^2)[z^2(\nu,\gamma^2)-\delta^2]-\delta^2 z(\nu,\gamma^2)=\frac{1}{2}[z^3(\nu,\gamma^2)-3\delta^2 z(\nu,\gamma^2)] \end{aligned}$$

and

$$\gamma^2\gamma^2 : \frac{1}{4}[z^2(\nu,\gamma^2)-\delta^2]^2-\delta^2[z^2(\nu,\gamma^2)-\delta^2]=\frac{1}{4}[z^4(\nu,\gamma^2)-6\delta^2 z^2(\nu,\gamma^2)+3\delta^4].$$

Under the null of correct specification, the expected value of these three terms should be zero. However, the expected value of the first term will also be zero under misspecification, so the test should only be based on the other two terms, which coincide with the third- and fourth-order Hermite polynomials of $z(\nu,\gamma^2)$, as claimed.

B.2.2 The bivariate case

The contribution of $\mathbf{x}=(x_1,x_2)'$ to the log-likelihood function is

$$-\ln 2\pi+\frac{1}{2}\ln|\boldsymbol{\Delta}|-\frac{1}{2}\boldsymbol{\varepsilon}'(\boldsymbol{\nu})\boldsymbol{\Delta}\boldsymbol{\varepsilon}(\boldsymbol{\nu}),$$

where $\boldsymbol{\nu}=(\nu_1,\nu_2)'$ and $vech(\boldsymbol{\Delta})=(\delta_{11},\delta_{12},\delta_{22})$.

If we suppress the dependence on the means for notational simplicity, the scores of this component with respect to the vector of mean parameters are

$$\mathbf{s}_{\boldsymbol{\nu}}(\mathbf{x};\boldsymbol{\nu},\boldsymbol{\gamma})=\begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{12} & \delta_{22} \end{pmatrix}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}=\begin{pmatrix} \delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2 \\ \delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2 \end{pmatrix},$$

which coincide with the $H_{10}(\boldsymbol{\varepsilon},\boldsymbol{\Delta})$ and $H_{01}(\boldsymbol{\varepsilon},\boldsymbol{\Delta})$ bivariate Hermite polynomials of $\boldsymbol{\varepsilon}$ in Barndorff-Nielsen and Petersen (1979).

Similarly, the scores with respect to the covariance matrix parameters $\boldsymbol{\gamma}=(\gamma_{11},\gamma_{12},\gamma_{22})'$ are given by one half of the product of the transpose of the duplication matrix

$$D_2'=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

times

$$vec\left[\begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{12} & \delta_{22} \end{pmatrix}\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}\begin{pmatrix} \varepsilon_1 & \varepsilon_2 \end{pmatrix}\begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{12} & \delta_{22} \end{pmatrix}-\begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{12} & \delta_{22} \end{pmatrix}\right]$$

$$=\left[\begin{array}{c} \delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11} \\ \delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12} \\ \delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12} \\ \delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22} \end{array}\right],$$

which coincide with the $H_{20}(\boldsymbol{\varepsilon},\boldsymbol{\Delta})$, $H_{11}(\boldsymbol{\varepsilon},\boldsymbol{\Delta})$ and $H_{02}(\boldsymbol{\varepsilon},\boldsymbol{\Delta})$ bivariate Hermite polynomials of $\boldsymbol{\varepsilon}$ in Barndorff-Nielsen and Petersen (1979). Therefore, the $\boldsymbol{\nu\nu}$ term of the sum of the outer product of the score and the Hessian matrix are identical to these polynomials.

In turn, the $\boldsymbol{\gamma\nu}$ term is one half the transpose of the duplication matrix times

$$\left[\begin{array}{c} (\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ (\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \end{array}\right.$$

$$\left.\begin{array}{c} (\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \\ (\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \end{array}\right]$$

$$-2\begin{bmatrix} \delta_{11}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) & \delta_{12}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ \delta_{12}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) & \delta_{22}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ \delta_{11}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) & \delta_{12}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \\ \delta_{12}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) & \delta_{22}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \end{bmatrix},$$

which reduces to

$$\left[\begin{array}{c} (\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ 2\left(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2\right) \\ (\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \end{array}\right.$$

$$\left.\begin{array}{c} (\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \\ 2\left(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2\right) \\ (\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \end{array}\right]$$

$$-2\begin{bmatrix} \delta_{11}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) & \delta_{12}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ 2\delta_{11}\delta_{12}\varepsilon_1+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_2 & (\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1+2\delta_{22}\delta_{12}\varepsilon_2 \\ \delta_{12}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) & \delta_{22}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \end{bmatrix}$$

$$=\left[\begin{array}{c} (\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2)-2\delta_{11}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ 2(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2)-2(2\delta_{11}\delta_{12}\varepsilon_1+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_2) \\ (\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2)-2\delta_{12}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \end{array}\right.$$

$$\left.\begin{array}{c} (\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2)-2\delta_{12}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ 2(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2-\delta_{12})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2)-2((\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1+2\delta_{22}\delta_{12}\varepsilon_2) \\ (\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2-\delta_{22})(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2)-2\delta_{22}(\delta_{12}\varepsilon_1+\delta_{22}\varepsilon_2) \end{array}\right]$$

It is tedious but trivial to see that the (2,1) and (2,2) elements are twice as big as the (1,2) and (3,1) ones, respectively. Therefore, the number of different elements coincides with the number of different third moments, which is $M(M+1)(M+2)/6 = 4$ in the bivariate case. Those four terms are

$$\begin{aligned} &(\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11})(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2)-2\delta_{11}(\delta_{11}\varepsilon_1+\delta_{12}\varepsilon_2) \\ = \; &\delta_{11}^3\varepsilon_1^3+3\delta_{11}^2\delta_{12}\varepsilon_1^2\varepsilon_2+3\delta_{11}\delta_{12}^2\varepsilon_2^2\varepsilon_1+\delta_{12}^3\varepsilon_2^3-3\delta_{11}^2\varepsilon_1-3\delta_{11}\delta_{12}\varepsilon_2=H_{30}(\boldsymbol{\varepsilon},\boldsymbol{\Delta}), \end{aligned}$$

$$(\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11})(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) - 2\delta_{12}(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2)$$
$$= \delta_{11}^2\delta_{12}\varepsilon_1^3 + \left(\delta_{22}\delta_{11}^2 + 2\delta_{11}\delta_{12}^2\right)\varepsilon_1^2\varepsilon_2 + (\delta_{12}^3 + 2\delta_{11}\delta_{22}\delta_{12})\varepsilon_2^2\varepsilon_1 + \delta_{22}\delta_{12}^2\varepsilon_2^3$$
$$-3\delta_{11}\delta_{12}\varepsilon_1 - (2\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_2 = H_{21}(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}),$$

$$(\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) - 2\delta_{12}(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2)$$
$$= \delta_{22}^2\delta_{12}\varepsilon_2^3 + \left(\delta_{11}\delta_{22}^2 + 2\delta_{22}\delta_{12}^2\right)\varepsilon_2^2\varepsilon_1 + \left(\delta_{12}^3 + 2\delta_{11}\delta_{22}\delta_{12}\right)\varepsilon_1^2\varepsilon_2 + \delta_{11}\delta_{12}^2\varepsilon_1^3$$
$$-(2\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1 - 3\delta_{22}\delta_{12}\varepsilon_2 = H_{12}(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}),$$

and

$$(\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) - 2\delta_{22}(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2)$$
$$= \delta_{22}^3\varepsilon_2^3 + 3\delta_{22}^2\delta_{12}\varepsilon_2^2\varepsilon_1 + 3\delta_{22}\delta_{12}^2\varepsilon_1^2\varepsilon_2 + \delta_{12}^3\varepsilon_1^3 - 3\delta_{22}\delta_{12}\varepsilon_1 - 3\delta_{22}^2\varepsilon_2 = H_{03}(\boldsymbol{\varepsilon}, \boldsymbol{\Delta}),$$

which coincide with the four different bivariate Hermite polynomials of order three in Barndorff-Nielsen and Petersen (1979), as expected.

Finally, the $\boldsymbol{\gamma\gamma}$ term of the outer product of the score is one quarter of

$$\begin{bmatrix} \delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11} \\ 2(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12}) \\ \delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22} \end{bmatrix}$$
$$\times \begin{bmatrix} \delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11} \\ 2(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12}) \\ \delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22} \end{bmatrix}'$$
$$= \begin{bmatrix} (\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11})^2 \\ 2(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})(\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11}) \\ (\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})(\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11}) \end{bmatrix.$$
$$\begin{matrix} 2(\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11})(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12}) \\ 4(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})^2 \\ 2(\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12}) \end{matrix}$$
$$\left. \begin{matrix} (\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11})(\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22}) \\ 2(\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})(\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22}) \\ (\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})^2 \end{matrix} \right].$$

To obtain the Hessian, we need the following matrix

$$\begin{bmatrix}
2\delta_{11}(\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2)-\delta_{11}^2 & 2\delta_{11}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{11}\delta_{12} & 2\delta_{12}(\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2)-\delta_{12}\delta_{11} & 2\delta_{12}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{12}^2 \\
2\delta_{11}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{11}\delta_{12} & 2\delta_{11}(\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2)-\delta_{11}\delta_{22} & 2\delta_{12}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{12}^2 & 2\delta_{12}(\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2)-\delta_{12}\delta_{22} \\
2\delta_{12}(\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2)-\delta_{12}\delta_{11} & 2\delta_{12}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{12}^2 & 2\delta_{22}(\delta_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2)-\delta_{22}\delta_{11} & 2\delta_{22}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{22}\delta_{12} \\
2\delta_{12}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{12}^2 & 2\delta_{12}(\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2)-\delta_{12}\delta_{22} & 2\delta_{22}(\delta_{11}\delta_{12}\varepsilon_1^2+(\delta_{12}^2+\delta_{11}\delta_{22})\varepsilon_1\varepsilon_2+\delta_{22}\delta_{12}\varepsilon_2^2)-\delta_{22}\delta_{12} & 2\delta_{22}(\delta_{12}^2\varepsilon_1^2+2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2+\delta_{22}^2\varepsilon_2^2)-\delta_{22}^2
\end{bmatrix}$$

which postmultiplied by the duplication matrix and premultiplied by its transpose yields

$$\begin{bmatrix}
\delta_{11}\left(2\varepsilon_1^2\delta_{11}^2+4\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}+2\varepsilon_2^2\delta_{12}^2-\delta_{11}\right) & 4\varepsilon_1^2\delta_{11}^2\delta_{12}+2\delta_{22}\varepsilon_1\varepsilon_2\delta_{11}^2+6\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}^2+2\delta_{22}\varepsilon_2^2\delta_{11}\delta_{12}+2\varepsilon_2^2\delta_{12}^3-2\delta_{11}\delta_{12} & \delta_{12}\left(2\delta_{11}\varepsilon_1^2\delta_{12}+2\varepsilon_1\varepsilon_2\delta_{12}^2+2\delta_{11}\delta_{22}\varepsilon_1\varepsilon_2+2\delta_{22}\varepsilon_2^2\delta_{12}-\delta_{12}\right) \\
4\varepsilon_1^2\delta_{11}^2\delta_{12}+2\delta_{22}\varepsilon_1\varepsilon_2\delta_{11}^2+6\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}^2+2\delta_{22}\varepsilon_2^2\delta_{11}\delta_{12}+2\varepsilon_2^2\delta_{12}^3-2\delta_{11}\delta_{12} & 2\varepsilon_1^2\delta_{11}^2\delta_{22}+6\varepsilon_1^2\delta_{11}\delta_{12}^2+12\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}\delta_{22}+4\varepsilon_1\varepsilon_2\delta_{12}^3+2\varepsilon_2^2\delta_{11}\delta_{22}^2+6\varepsilon_2^2\delta_{12}^2\delta_{22}-2\delta_{11}\delta_{22}-2\delta_{12}^2 & 2\varepsilon_1^2\delta_{12}^3+2\delta_{11}\varepsilon_1^2\delta_{12}\delta_{22}+6\varepsilon_1\varepsilon_2\delta_{12}^2\delta_{22}+2\delta_{11}\varepsilon_1\varepsilon_2\delta_{22}^2+4\varepsilon_2^2\delta_{12}\delta_{22}^2-2\delta_{12}\delta_{22} \\
\delta_{12}\left(2\delta_{11}\varepsilon_1^2\delta_{12}+2\varepsilon_1\varepsilon_2\delta_{12}^2+2\delta_{11}\delta_{22}\varepsilon_1\varepsilon_2+2\delta_{22}\varepsilon_2^2\delta_{12}-\delta_{12}\right) & 2\varepsilon_1^2\delta_{12}^3+2\delta_{11}\varepsilon_1^2\delta_{12}\delta_{22}+6\varepsilon_1\varepsilon_2\delta_{12}^2\delta_{22}+2\delta_{11}\varepsilon_1\varepsilon_2\delta_{22}^2+4\varepsilon_2^2\delta_{12}\delta_{22}^2-2\delta_{12}\delta_{22} & \delta_{22}\left(2\varepsilon_1^2\delta_{12}^2+4\varepsilon_1\varepsilon_2\delta_{12}\delta_{22}+2\varepsilon_2^2\delta_{22}^2-\delta_{22}\right)
\end{bmatrix}$$

If we subtract twice this matrix from the compressed outer product of the score we end up with a 3×3 matrix with the following elements

$$\begin{aligned}
&\mathbf{(1,1)}:\varepsilon_1^4\delta_{11}^4+4\varepsilon_1^3\varepsilon_2\delta_{11}^3\delta_{12}+6\varepsilon_1^2\varepsilon_2^2\delta_{11}^2\delta_{12}^2-6\varepsilon_1^2\delta_{11}^3+4\varepsilon_1\varepsilon_2^3\delta_{11}\delta_{12}^3\\
&\quad-12\varepsilon_1\varepsilon_2\delta_{11}^2\delta_{12}+\varepsilon_2^4\delta_{12}^4-6\varepsilon_2^2\delta_{11}\delta_{12}^2+3\delta_{11}^2,\\
&\mathbf{(2,1)}:2\varepsilon_1^4\delta_{11}^3\delta_{12}+2\delta_{22}\varepsilon_1^3\varepsilon_2\delta_{11}^3+6\varepsilon_1^3\varepsilon_2\delta_{11}^2\delta_{12}^2+6\delta_{22}\varepsilon_1^2\varepsilon_2^2\delta_{11}^2\delta_{12}+6\varepsilon_1^2\varepsilon_2^2\delta_{11}\delta_{12}^3\\
&\quad-12\varepsilon_1^2\delta_{11}^2\delta_{12}+6\delta_{22}\varepsilon_1\varepsilon_2^3\delta_{11}\delta_{12}^2+2\varepsilon_1\varepsilon_2^3\delta_{12}^4-6\delta_{22}\varepsilon_1\varepsilon_2\delta_{11}^2\\
&\quad-18\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}^2+2\delta_{22}\varepsilon_2^4\delta_{12}^3-6\delta_{22}\varepsilon_2^2\delta_{11}\delta_{12}-6\varepsilon_2^2\delta_{12}^3+6\delta_{11}\delta_{12},\\
&\mathbf{(3,1)}:\varepsilon_1^4\delta_{11}^2\delta_{12}^2+2\varepsilon_1^3\varepsilon_2\delta_{11}^2\delta_{12}\delta_{22}+2\varepsilon_1^3\varepsilon_2\delta_{11}\delta_{12}^3+\varepsilon_1^2\varepsilon_2^2\delta_{11}^2\delta_{22}^2+4\varepsilon_1^2\varepsilon_2^2\delta_{11}\delta_{12}^2\delta_{22}+\varepsilon_1^2\varepsilon_2^2\delta_{12}^4\\
&\quad-\varepsilon_1^2\delta_{11}^2\delta_{22}-5\varepsilon_1^2\delta_{11}\delta_{12}^2+2\varepsilon_1\varepsilon_2^3\delta_{11}\delta_{12}\delta_{22}^2+2\varepsilon_1\varepsilon_2^3\delta_{12}^3\delta_{22}-8\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}\delta_{22}\\
&\quad-4\varepsilon_1\varepsilon_2\delta_{12}^3+\varepsilon_2^4\delta_{12}^2\delta_{22}^2-\varepsilon_2^2\delta_{11}\delta_{22}^2-5\varepsilon_2^2\delta_{12}^2\delta_{22}+\delta_{11}\delta_{22}+2\delta_{12}^2,
\end{aligned}$$

$$\begin{aligned}
&\mathbf{(1,2)}:2\varepsilon_1^4\delta_{11}^3\delta_{12}+2\delta_{22}\varepsilon_1^3\varepsilon_2\delta_{11}^3+6\varepsilon_1^3\varepsilon_2\delta_{11}^2\delta_{12}^2+6\delta_{22}\varepsilon_1^2\varepsilon_2^2\delta_{11}^2\delta_{12}\\
&\quad+6\varepsilon_1^2\varepsilon_2^2\delta_{11}\delta_{12}^3-12\varepsilon_1^2\delta_{11}^2\delta_{12}+6\delta_{22}\varepsilon_1\varepsilon_2^3\delta_{11}\delta_{12}^2+2\varepsilon_1\varepsilon_2^3\delta_{12}^4-6\delta_{22}\varepsilon_1\varepsilon_2\delta_{11}^2\\
&\quad-18\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}^2+2\delta_{22}\varepsilon_2^4\delta_{12}^3-6\delta_{22}\varepsilon_2^2\delta_{11}\delta_{12}-6\varepsilon_2^2\delta_{12}^3+6\delta_{11}\delta_{12},\\
&\mathbf{(2,2)}:4\varepsilon_1^4\delta_{11}^2\delta_{12}^2+8\varepsilon_1^3\varepsilon_2\delta_{11}^2\delta_{12}\delta_{22}+8\varepsilon_1^3\varepsilon_2\delta_{11}\delta_{12}^3+4\varepsilon_1^2\varepsilon_2^2\delta_{11}^2\delta_{22}^2+16\varepsilon_1^2\varepsilon_2^2\delta_{11}\delta_{12}^2\delta_{22}+4\varepsilon_1^2\varepsilon_2^2\delta_{12}^4\\
&\quad-4\varepsilon_1^2\delta_{11}^2\delta_{22}-20\varepsilon_1^2\delta_{11}\delta_{12}^2+8\varepsilon_1\varepsilon_2^3\delta_{11}\delta_{12}\delta_{22}^2+8\varepsilon_1\varepsilon_2^3\delta_{12}^3\delta_{22}-32\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}\delta_{22}\\
&\quad-16\varepsilon_1\varepsilon_2\delta_{12}^3+4\varepsilon_2^4\delta_{12}^2\delta_{22}^2-4\varepsilon_2^2\delta_{11}\delta_{22}^2-20\varepsilon_2^2\delta_{12}^2\delta_{22}+4\delta_{11}\delta_{22}+8\delta_{12}^2,\\
&\mathbf{(3,2)}:2\delta_{11}\varepsilon_1^4\delta_{12}^3+2\varepsilon_1^3\varepsilon_2\delta_{12}^4+6\delta_{11}\varepsilon_1^3\varepsilon_2\delta_{12}^2\delta_{22}+6\varepsilon_1^2\varepsilon_2^2\delta_{12}^3\delta_{22}\\
&\quad+6\delta_{11}\varepsilon_1^2\varepsilon_2^2\delta_{12}\delta_{22}^2-6\varepsilon_1^2\delta_{12}^3-6\delta_{11}\varepsilon_1^2\delta_{12}\delta_{22}+6\varepsilon_1\varepsilon_2^3\delta_{12}^2\delta_{22}^2+2\delta_{11}\varepsilon_1\varepsilon_2^3\delta_{22}^3\\
&\quad-18\varepsilon_1\varepsilon_2\delta_{12}^2\delta_{22}-6\delta_{11}\varepsilon_1\varepsilon_2\delta_{22}^2+2\varepsilon_2^4\delta_{12}\delta_{22}^3-12\varepsilon_2^2\delta_{12}\delta_{22}^2+6\delta_{12}\delta_{22},
\end{aligned}$$

$$\begin{gathered}(\mathbf{1},\mathbf{3}):\varepsilon_1^4\delta_{11}^2\delta_{12}^2+2\varepsilon_1^3\varepsilon_2\delta_{11}^2\delta_{12}\delta_{22}+2\varepsilon_1^3\varepsilon_2\delta_{11}\delta_{12}^3+\varepsilon_1^2\varepsilon_2^2\delta_{11}^2\delta_{22}^2\\ +4\varepsilon_1^2\varepsilon_2^2\delta_{11}\delta_{12}^2\delta_{22}+\varepsilon_1^2\varepsilon_2^2\delta_{12}^4-\varepsilon_1^2\delta_{11}^2\delta_{22}-5\varepsilon_1^2\delta_{11}\delta_{12}^2+2\varepsilon_1\varepsilon_2^3\delta_{11}\delta_{12}\delta_{22}^2+2\varepsilon_1\varepsilon_2^3\delta_{12}^3\delta_{22}\\ -8\varepsilon_1\varepsilon_2\delta_{11}\delta_{12}\delta_{22}-4\varepsilon_1\varepsilon_2\delta_{12}^3+\varepsilon_2^4\delta_{12}^2\delta_{22}^2-\varepsilon_2^2\delta_{11}\delta_{22}^2-5\varepsilon_2^2\delta_{12}^2\delta_{22}+\delta_{11}\delta_{22}+2\delta_{12}^2,\\ (\mathbf{2},\mathbf{3}):2\delta_{11}\varepsilon_1^4\delta_{12}^3+2\varepsilon_1^3\varepsilon_2\delta_{12}^4+6\delta_{11}\varepsilon_1^3\varepsilon_2\delta_{12}^2\delta_{22}+6\varepsilon_1^2\varepsilon_2^2\delta_{12}^3\delta_{22}+6\delta_{11}\varepsilon_1^2\varepsilon_2^2\delta_{12}\delta_{22}^2\\ -6\varepsilon_1^2\delta_{12}^3-6\delta_{11}\varepsilon_1^2\delta_{12}\delta_{22}+6\varepsilon_1\varepsilon_2^3\delta_{12}^2\delta_{22}^2+2\delta_{11}\varepsilon_1\varepsilon_2^3\delta_{22}^3-18\varepsilon_1\varepsilon_2\delta_{12}^2\delta_{22}\\ -6\delta_{11}\varepsilon_1\varepsilon_2\delta_{22}^2+2\varepsilon_2^4\delta_{12}\delta_{22}^3-12\varepsilon_2^2\delta_{12}\delta_{22}^2+6\delta_{12}\delta_{22}\text{ and}\\ (\mathbf{3},\mathbf{3}):\varepsilon_1^4\delta_{12}^4+4\varepsilon_1^3\varepsilon_2\delta_{12}^3\delta_{22}+6\varepsilon_1^2\varepsilon_2^2\delta_{12}^2\delta_{22}^2-6\varepsilon_1^2\delta_{12}^2\delta_{22}+4\varepsilon_1\varepsilon_2^3\delta_{12}\delta_{22}^3\\ -12\varepsilon_1\varepsilon_2\delta_{12}\delta_{22}^2+\varepsilon_2^4\delta_{22}^4-6\varepsilon_2^2\delta_{22}^3+3\delta_{22}^2.\end{gathered}$$

Once again, it is tedious but straightforward to prove that the elements (2,1), (3,1) and (3,2) are equal to the elements (1,2), (1,3) and (2,3), respectively. In addition, the (2,2) element is four times the (3,1) and (1,3) ones. Therefore, the number of different elements coincides with the number of different fourth moments, which is $M(M+1)(M+2)(M+3)/24=5$ in the bivariate case. Those five terms are

$$\begin{gathered}\delta_{11}^4\varepsilon_1^4+4\delta_{11}^3\delta_{12}\varepsilon_1^3\varepsilon_2+6\delta_{11}^2\delta_{12}^2\varepsilon_1^2\varepsilon_2^2+4\delta_{11}\delta_{12}^3\varepsilon_1\varepsilon_2^3+\delta_{12}^4\varepsilon_2^4\\ -6\delta_{11}^3\varepsilon_1^2-12\delta_{11}^2\delta_{12}\varepsilon_1\varepsilon_2-6\delta_{11}\delta_{12}^2\varepsilon_2^2+3\delta_{11}^2=H_{40}(\boldsymbol{\varepsilon},\boldsymbol{\Delta}),\end{gathered}$$

$$\begin{gathered}2\delta_{11}^3\delta_{12}\varepsilon_1^4+2(\delta_{22}\delta_{11}^3+3\delta_{11}^2\delta_{12}^2)\varepsilon_1^3\varepsilon_2+6(\delta_{22}\delta_{11}^2\delta_{12}+\delta_{11}\delta_{12}^3)\varepsilon_1^2\varepsilon_2^2\\ +2(3\delta_{22}\delta_{11}\delta_{12}^2+\delta_{12}^4)\varepsilon_1\varepsilon_2^3+2\delta_{22}\delta_{12}^3\varepsilon_2^4\\ -12\delta_{11}^2\delta_{12}\varepsilon_1^2-6(\delta_{22}\delta_{11}^2+3\delta_{11}\delta_{12}^2)\varepsilon_1\varepsilon_2-6(\delta_{22}\delta_{11}\delta_{12}+\delta_{12}^3)\varepsilon_2^2+6\delta_{11}\delta_{12}=2H_{31}(\boldsymbol{\varepsilon},\boldsymbol{\Delta}),\end{gathered}$$

$$\begin{gathered}\delta_{11}^2\delta_{12}^2\varepsilon_1^4+2\left(\delta_{22}\delta_{11}^2\delta_{12}+\delta_{11}\delta_{12}^3\right)\varepsilon_2\varepsilon_1^3+\left(\delta_{11}^2\delta_{22}^2+4\delta_{11}\delta_{12}^2\delta_{22}+\delta_{12}^4\right)\varepsilon_2^2\varepsilon_1^2\\ +2\left(\delta_{12}^3\delta_{22}+\delta_{11}\delta_{12}\delta_{22}^2\right)\varepsilon_2^3\varepsilon_1+\varepsilon_2^4\delta_{12}^2\delta_{22}^2-\left(\delta_{11}^2\delta_{22}+5\delta_{11}\delta_{12}^2\right)\varepsilon_1^2\\ -4\left(\delta_{12}^3+2\delta_{11}\delta_{12}\delta_{22}\right)\varepsilon_1\varepsilon_2-(5\delta_{12}^2\delta_{22}+\delta_{11}\delta_{22}^2)\varepsilon_2^2+\left(2\delta_{12}^2+\delta_{11}\delta_{22}\right)=H_{22}(\boldsymbol{\varepsilon},\boldsymbol{\Delta}),\end{gathered}$$

$$\begin{gathered}2\delta_{11}\delta_{12}^3\varepsilon_1^4+2\left(\delta_{12}^4+3\delta_{11}\delta_{22}\delta_{12}^2\right)\varepsilon_1^3\varepsilon_2+6\left(\delta_{12}^3\delta_{22}+\delta_{11}\delta_{12}\delta_{22}^2\right)\varepsilon_1^2\varepsilon_2^2\\ +2\left(3\delta_{12}^2\delta_{22}^2+\delta_{11}\delta_{22}^3\right)\varepsilon_2^3\varepsilon_1+2\delta_{12}\delta_{22}^3\varepsilon_2^4-6\left(\delta_{12}^3+\delta_{11}\delta_{12}\delta_{22}\right)\varepsilon_1^2\\ -6\left(3\delta_{12}^2\delta_{22}+\delta_{11}\delta_{22}^2\right)\varepsilon_1\varepsilon_2-12\delta_{12}\delta_{22}^2\varepsilon_2^2+6\delta_{12}\delta_{22}=2H_{13}(\boldsymbol{\varepsilon},\boldsymbol{\Delta})\end{gathered}$$

and

$$\begin{gathered}\delta_{12}^4\varepsilon_1^4+4\delta_{12}^3\delta_{22}\varepsilon_1^3\varepsilon_2+6\delta_{12}^2\delta_{22}^2\varepsilon_1^2\varepsilon_2^2+4\delta_{12}\delta_{22}^3\varepsilon_1\varepsilon_2^3+\delta_{22}^4\varepsilon_2^4\\ -6\delta_{12}^2\delta_{22}\varepsilon_1^2-12\delta_{12}\delta_{22}^2\varepsilon_1\varepsilon_2-6\delta_{22}^3\varepsilon_2^2+3\delta_{22}^2=H_{04}(\boldsymbol{\varepsilon},\boldsymbol{\Delta}),\end{gathered}$$

which are (multiples of) the five different bivariate Hermite polynomials of order four in Barndorff-Nielsen and Petersen (1979), as expected.

B.3 Alternative distributions

For the multivariate skew normal distribution, we use its canonical representation, choosing 0.83, 1.30 and -1.35 for the location, scale and skew, respectively, of the first component of the random vector, which yield values of $-3/4$ and 3.60 for its skewness and kurtosis coefficients (see Figure 2.2 in Azzalini and Capitanio (2014) for the feasible skewness-kurtosis combinations). In contrast, the remaining $M - 1$ components are drawn from independent univariate standard normals.

In the case of the multivariate asymmetric Student t, we choose $\eta = 0.042$ and $\mathbf{b} = (-0.91, \mathbf{0}')'$, which yield values of $-3/4$ and 4.5 for the skewness and kurtosis coefficients of the first element (see Proposition 1 in Mencía and Sentana (2009) for details on how to obtain a random vector whose mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively). Finally, for the discrete mixture of two normal vectors, we fix their means to $(1 - \lambda)\boldsymbol{\upsilon}$ and $-\lambda\boldsymbol{\upsilon}$, where $\lambda = 1/4$ is the probability of the first Gaussian vector and $\boldsymbol{\upsilon} = (-.57, \mathbf{0}')'$, and their covariance matrices to

$$\begin{aligned}\boldsymbol{\Omega}_1 &= \frac{1}{\lambda + \varkappa(1-\lambda)}\left[\mathbf{I}_M - \boldsymbol{\upsilon}\boldsymbol{\upsilon}'(1-\lambda)\lambda\right] \text{ and} \\ \boldsymbol{\Omega}_2 &= \varkappa\boldsymbol{\Omega}_1\end{aligned}$$

with $\varkappa = .51$, so as to achieve the same skewness and kurtosis coefficients for the first variable as in the case of the asymmetric Student t.

Table 1: Joint normality tests: Size (asymptotic critical values)

Monte Carlo rejection rates at the 5% significance level

Panel A: Information matrix tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	4.79	4.98	5.00	*5*	5.44	5.70	5.36	*9*	6.49	6.33	5.59
4	*20*	4.81	5.03	5.05	*35*	8.51	8.29	6.53	*55*	9.30	8.61	6.61
8	*120*	4.11	4.96	4.95	*330*	10.01	11.33	8.19	*450*	10.55	11.56	8.23
16	*816*	2.23	4.35	4.88	*3876*	6.07	13.14	11.14	*4692*	6.44	13.39	11.18

Panel B: Kilian and Demiroglu (2000) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*2*	4.59	4.88	4.97	*2*	3.89	4.62	4.92	*4*	5.09	5.31	5.18
4	*4*	4.56	4.91	5.00	*4*	4.76	5.29	5.22	*8*	5.91	5.92	5.47
8	*8*	4.43	4.89	4.99	*8*	5.51	5.91	5.40	*16*	6.44	6.40	5.59
16	*16*	4.15	4.78	4.97	*16*	5.93	6.35	5.64	*32*	6.58	6.61	5.69

Panel C: Mardia (1970) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	4.79	4.98	5.00	*1*	2.95	4.33	4.80	*5*	4.64	5.01	5.02
4	*20*	4.81	5.03	5.05	*1*	3.24	4.46	4.83	*21*	4.67	5.06	5.06
8	*120*	4.11	4.96	4.95	*1*	5.17	4.98	4.97	*121*	3.97	4.92	4.95
16	*816*	2.23	4.35	4.88	*1*	16.23	7.47	5.61	*817*	2.16	4.31	4.87

Notes: We generate 20,000 samples from a spherical Gaussian random vector. *df* denotes degrees of freedom.

Table 2: Joint normality tests: Size (simulated critical values)

Monte Carlo rejection rates at the 5% significance level

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
Panel A: Information matrix tests												
2	*4*	4.74	5.15	4.75	*5*	4.75	5.08	4.85	*9*	4.87	4.87	4.71
4	*20*	4.89	5.30	4.84	*35*	4.92	5.02	4.97	*55*	4.93	5.21	4.91
8	*120*	5.00	5.04	5.09	*330*	5.09	4.72	4.81	*450*	5.07	4.86	4.83
16	*816*	5.07	4.88	5.12	*3876*	4.94	4.88	4.96	*4692*	4.93	4.92	4.96
Panel B: Kilian and Demiroglu (2000) tests												
2	*2*	4.49	5.19	4.72	*2*	4.86	4.99	4.88	*4*	4.83	5.07	4.76
4	*4*	5.21	5.16	4.86	*4*	5.12	4.98	4.93	*8*	5.34	4.87	4.70
8	*8*	4.93	4.83	5.10	*8*	5.13	5.07	5.18	*16*	4.98	5.08	5.02
16	*16*	4.97	5.24	4.98	*16*	5.15	5.12	4.75	*32*	5.17	4.90	5.06
Panel C: Mardia (1970) tests												
2	*4*	4.74	5.15	4.75	*1*	4.86	4.88	4.66	*5*	4.75	4.96	4.74
4	*20*	4.89	5.30	4.84	*1*	4.95	4.81	5.05	*21*	4.91	5.37	4.86
8	*120*	5.00	5.04	5.09	*1*	5.09	5.03	5.04	*121*	5.01	5.05	5.00
16	*816*	5.07	4.88	5.12	*1*	4.73	5.15	5.09	*817*	5.08	4.88	5.12

Notes: We approximate the exact finite sample critical values with $R = 10^6$ replications from a spherical Gaussian random vector. We generate 20,000 additional samples to compute the rejection rates. *df* denotes degrees of freedom.

Table 3: Marginal, conditional and remainder components of the information matrix test: Size

Monte Carlo rejection rates at the 5% significance level, $N = 400$

	df	Asymptotic critical values	Simulated critical values
Panel A: Marginal			
Normality of regressor			
$H_3(x)$ (skewness)	*1*	4.90	4.98
$H_4(x)$ (kurtosis)	*1*	4.21	5.03
$H_3(x)$ & $H_4(x)$ (Jarque-Bera)	*2*	4.77	5.06
Panel B: Conditional bivariate			
Normality of residuals			
$H_3(u)$ (skewness)	*1*	5.08	5.22
$H_4(u)$ (kurtosis)	*1*	4.33	5.18
$H_3(u)$ & $H_4(u)$ (Jarque-Bera)	*2*	4.77	5.09
Heteroskedasticity			
$H_2(u)H_1(x)$ & $H_2(u)H_2(x)$	*2*	4.78	5.01
Asymmetry			
$H_3(u)H_1(x)$ (conditional asymmetry)	*1*	4.96	5.19
$H_3(u)$ & $H_3(u)H_1(x)$ (total asymmetry)	*2*	5.34	5.17
Total			
$H_3(u)$ & $H_2(u)H_1(x)$ & $H_4(u)$ $H_3(u)H_1(x)$ & $H_2(u)H_2(x)$	*5*	5.57	4.95
Panel C: The "rest"			
$H_1(u)H_2(x)$ & $H_1(u)H_3(x)$	*2*	5.29	5.20
Panel D: Joint bivariate			
All of them	*9*	6.48	5.04

Notes: We approximate the exact finite sample critical values with $R = 10^6$ replications from a spherical Gaussian random vector. We generate 20,000 additional samples to compute the rejection rates. *df* denotes degrees of freedom and u denotes the residual of the linear regression of x_2 onto a constant and x_1.

Table 4: Joint normality tests: Power (asymmetric Student t)

Monte Carlo rejection rates at the 5% significance level

Panel A: Information matrix tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	21.73	51.97	96.70	*5*	22.09	47.65	91.22	*9*	24.32	58.58	98.77
4	*20*	28.73	59.73	97.47	*35*	32.76	70.29	99.22	*55*	34.56	76.41	99.86
8	*120*	50.58	83.86	99.60	*330*	58.93	97.18	100.00	*450*	59.98	97.77	100.00
16	*816*	86.34	99.73	100.00	*3876*	91.36	100.00	100.00	*4692*	91.57	100.00	100.00

Panel B: Kilian and Demiroglu (2000) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*2*	19.77	47.73	94.64	*2*	20.04	43.54	88.98	*4*	22.07	55.30	98.11
4	*4*	19.43	45.39	91.74	*4*	22.16	53.62	96.80	*8*	23.55	61.53	99.30
8	*8*	20.06	44.55	89.09	*8*	23.94	67.33	99.71	*16*	25.29	71.09	99.90
16	*16*	20.26	46.33	87.50	*16*	25.25	81.41	100.00	*32*	26.63	82.94	100.00

Panel C: Mardia (1970) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	21.73	51.97	96.70	*1*	21.11	52.14	96.07	*5*	25.32	62.84	99.33
4	*20*	28.73	59.73	97.47	*1*	27.62	80.36	99.98	*21*	33.18	77.01	99.98
8	*120*	50.58	83.86	99.60	*1*	37.91	99.11	100.00	*121*	53.49	94.10	100.00
16	*816*	86.34	99.73	100.00	*1*	42.82	100.00	100.00	*817*	86.62	99.94	100.00

Notes: We approximate the exact finite sample critical values with $R = 10^6$ replications from a spherical Gaussian random vector. We generate 20,000 samples from the asymmetric Student t distribution with mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively. See Supplemental Appendix B.3 for details. *df* denotes degrees of freedom.

Table 5: Joint normality tests: Power (mixture of two normals)

Monte Carlo rejection rates at the 5% significance level

Panel A: Information matrix tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	21.72	50.48	95.93	*5*	22.51	50.02	94.40	*9*	24.88	60.93	99.12
4	*20*	30.94	61.10	97.18	*35*	36.22	76.02	99.76	*55*	38.14	81.83	99.97
8	*120*	57.87	88.54	99.73	*330*	67.91	99.28	100.00	*450*	69.36	99.48	100.00
16	*816*	92.03	99.97	100.00	*3876*	96.34	100.00	100.00	*4692*	96.36	100.00	100.00

Panel B: Kilian and Demiroglu (2000) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*2*	19.66	46.45	93.28	*2*	20.26	46.03	92.27	*4*	22.50	57.38	98.55
4	*4*	19.97	44.16	90.46	*4*	23.70	58.54	98.61	*8*	25.14	65.44	99.61
8	*8*	21.77	44.78	88.17	*8*	27.68	73.97	99.98	*16*	29.17	76.98	99.99
16	*16*	22.76	48.94	87.11	*16*	29.64	88.93	100.00	*32*	31.38	89.70	100.00

Panel C: Mardia (1970) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	21.72	50.48	95.93	*1*	21.58	56.96	97.95	*5*	25.80	65.22	99.56
4	*20*	30.94	61.10	97.18	*1*	31.15	87.31	100.00	*21*	36.20	82.19	99.98
8	*120*	57.87	88.54	99.73	*1*	46.20	99.80	100.00	*121*	61.21	97.47	100.00
16	*816*	92.03	99.97	100.00	*1*	54.43	100.00	100.00	*817*	92.26	100.00	100.00

Notes: We approximate the exact finite sample critical values with $R = 10^6$ replications from a spherical Gaussian random vector. We generate 20,000 samples from the two-component location-scale mixture of normals discussed by Mencía and Sentana (2009) with mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively. See Supplemental Appendix B.3 for details. *df* denotes degrees of freedom.

Table 6: Joint normality tests: Power (skew normal)

Monte Carlo rejection rates at the 5% significance level

Panel A: Information matrix tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	10.66	32.46	91.32	*5*	7.42	9.61	16.41	*9*	9.16	21.66	80.33
4	*20*	7.31	16.24	65.64	*35*	6.34	7.57	10.39	*55*	6.69	11.19	42.02
8	*120*	5.92	8.77	28.08	*330*	5.83	6.07	7.09	*450*	5.91	6.84	14.29
16	*816*	5.38	6.37	10.99	*3876*	5.46	5.56	6.31	*4692*	5.51	5.71	7.55

Panel B: Kilian and Demiroglu (2000) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*2*	13.36	40.89	95.77	*2*	7.79	10.69	19.41	*4*	11.19	31.23	90.88
4	*4*	9.98	31.55	91.00	*4*	6.74	9.19	15.93	*8*	8.42	22.70	81.85
8	*8*	7.86	23.30	82.48	*8*	6.35	7.69	12.61	*16*	7.12	16.43	68.88
16	*16*	6.44	16.98	69.04	*16*	5.50	6.75	10.24	*32*	5.96	12.31	53.49

Panel C: Mardia (1970) tests

	(co-)skewness				(co-)kurtosis				both			
$M \setminus N$	*df*	100	400	1,600	*df*	100	400	1,600	*df*	100	400	1,600
2	*4*	10.66	32.46	91.32	*1*	6.53	8.31	12.71	*5*	10.33	29.13	89.25
4	*20*	7.31	16.24	65.64	*1*	5.27	5.80	7.27	*21*	7.19	15.96	64.23
8	*120*	5.92	8.77	28.08	*1*	4.71	4.99	4.98	*121*	5.96	8.92	27.79
16	*816*	5.38	6.37	10.99	*1*	4.99	4.54	4.56	*817*	5.38	6.39	11.00

Notes: We approximate the exact finite sample critical values with $R = 10^6$ replications from a spherical Gaussian random vector. We generate 20,000 samples from the skew normal multivariate distribution in Azzalini and Dalla Valle (1996) with mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively. See Supplemental Appendix B.3 for details. *df* denotes degrees of freedom.

Table 7: Marginal, conditional and remainder components of the information matrix test: Power

Monte Carlo rejection rates at the 5% significance level, $N = 400$				
	df	Asymmetric Student t	Mixture of normals	Skew normal
Panel A: Marginal				
Normality of regressor				
$H_3(x)$ (skewness)	*1*	52.62	50.59	51.66
$H_4(x)$ (kurtosis)	*1*	36.81	37.40	12.70
$H_3(x)$ & $H_4(x)$ (Jarque-Bera)	*2*	54.33	54.80	41.88
Panel B: Conditional bivariate				
Normality of residuals				
$H_3(u)$ (skewness)	*1*	10.28	11.33	5.02
$H_4(u)$ (kurtosis)	*1*	24.34	28.19	5.08
$H_3(u)$ & $H_4(u)$ (Jarque-Bera)	*2*	23.16	26.41	5.05
Heteroskedasticity				
$H_2(u)H_1(x)$ & $H_2(u)H_2(x)$		33.44	33.77	5.29
Asymmetry				
$H_3(u)H_1(x)$ (conditional asymmetry)	*1*	15.33	15.68	4.98
$H_3(u)$ & $H_3(u)H_1(x)$ (total asymmetry)	*2*	16.00	16.89	4.96
Total				
$H_3(u)$ & $H_2(u)H_1(x)$ & $H_4(u)$ $H_3(u)H_1(x)$ & $H_2(u)H_2(x)$	*5*	37.20	40.65	5.27
Panel C: The "rest"				
$H_1(u)H_2(x)$ & $H_1(u)H_3(x)$	*2*	19.82	17.16	7.07
Panel D: Joint bivariate				
All of them	*9*	58.60	61.48	22.25

Notes: We approximate the exact finite sample critical values with $R = 10^6$ replications from a spherical Gaussian random vector. We generate 20,000 samples from three multivariate non-Gaussian distributions whose mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_M$, respectively: the asymmetric Student t distribution and the two-component location-scale mixture of normals discussed by Mencía and Sentana (2009), and the skew normal multivariate distribution in Azzalini and Dalla Valle (1996). See Supplemental Appendix B.3 for details. *df* denotes degrees of freedom and u denotes the residual of the linear regression of x_2 onto a constant and x_1.

Table 8: Testing for normality of the distribution of (log) city sizes and their growth rates

	Number of states that reject the null at the 5% significance level
Panel A: Marginal	
Normality	
$H_3(x_1)$ (skewness)	34
$H_4(x_1)$ (kurtosis)	19
$H_3(x_1)$ & $H_4(x_1)$ (Jarque-Bera)	37
Panel B: Conditional of x_2 given x_1	
Normality of residuals	
$H_3(u)$ (skewness)	44
$H_4(u)$ (kurtosis)	47
$H_3(u)$ & $H_4(u)$ (Jarque-Bera)	47
Heteroskedasticity	
$H_2(u)H_1(x_1)$ & $H_2(u)H_2(x_1)$	43
Asymmetry	
$H_3(u)H_1(x_1)$ (conditional asymmetry)	41
$H_3(u)$ & $H_3(u)H_1(x_1)$ (total asymmetry)	47
Total	
$H_3(u)$ & $H_2(u)H_1(x_1)$ & $H_4(u)$ & $H_3(u)H_1(x_1)$ & $H_2(u)H_2(x_1)$	47
Panel C: The "rest"	
$H_1(u)H_2(x_1)$ & $H_1(u)H_3(x_1)$	20
Panel D: Joint bivariate	
$H_3(x_1)$ & $H_3(u)$ & $H_2(u)H_1(x_1)$ & $H_4(x_1)$ & $H_4(u)$ & $H_3(u)H_1(x_1)$ & $H_2(u)H_2(x_1)$	47

Table 8: Testing for normality of the distribution of (log) city sizes and their growth rates (cont.)

	Number of states that reject the null at the 5% significance level
Panel E: Conditional of x_3 given x_1 and x_2	
Normality of residuals	
$H_3(v)$ (skewness)	39
$H_4(v)$ (kurtosis)	47
$H_3(v)$ & $H_4(v)$ (Jarque-Bera)	47
Heteroskedasticity	
$H_2(v)H_1(u)$ & $H_2(v)H_1(x_1)$ & $H_2(v)H_2(u)$ & $H_2(u)H_1(v)H_1(x_1)$ & $H_2(v)H_2(x_1)$	47
Asymmetry	
$H_3(v)H_1(u)$ & $H_3(v)H_1(x_1)$ (conditional asymmetry)	45
$H_3(v)$ & $H_3(v)H_1(u)$ & $H_3(v)H_1(x_1)$ (total asymmetry)	45
Total	
$H_3(v)$ & $H_2(v)H_1(u)$ & $H_2(v)H_1(x_1)$ & $H_4(v)$ & $H_3(v)H_1(u)$ & $H_3(v)H_1(x_1)$ & $H_2(v)H_2(u)$ & $H_2(u)H_1(v)H_1(x_1)$ & $H_2(v)H_2(x_1)$	47
Panel F: The "rest"	
$H_1(v)H_2(u)$ & $H_1(v)H_1(u)H_1(x_1)$ & $H_1(v)H_2(x_1)$ & $H_1(v)H_3(u)$ & $H_1(v)H_2(u)H_1(x_1)$ & $H_1(v)H_1(u)H_2(x_1)$ & $H_1(v)H_3(x_1)$	44
Panel G: Joint trivariate	
All of them	48

Notes: Samples consist on (log) city sizes in 2000 (x_1) and their (continuously compounded) growth rates between 2000 and 2010 (x_2), and 2010 and 2020 (x_3) for each of the 48 contiguous US states: 23,830 matched cities in the three censuses with a population of at least one in both years; see Section 5 for details. u and v denote the residual of the linear regression of x_2 onto a constant and x_1, and x_3 onto a constant, x_1 and x_2, respectively. See Section 3 for a detailed description of the test statistics.

Figure 1: Distribution of (log) city sizes and their growth rates

Figure 1a: (log) city sizes in 2000 and growth rates between 2000 and 2010

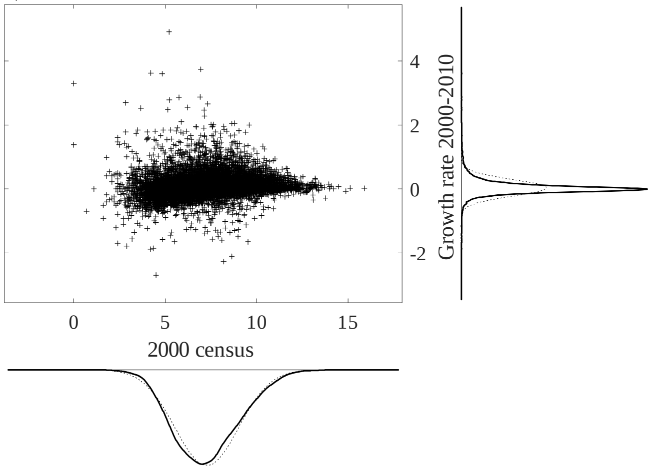


Figure1b: (log) city sizes in 2000 and growth rates between 2010 and 2020

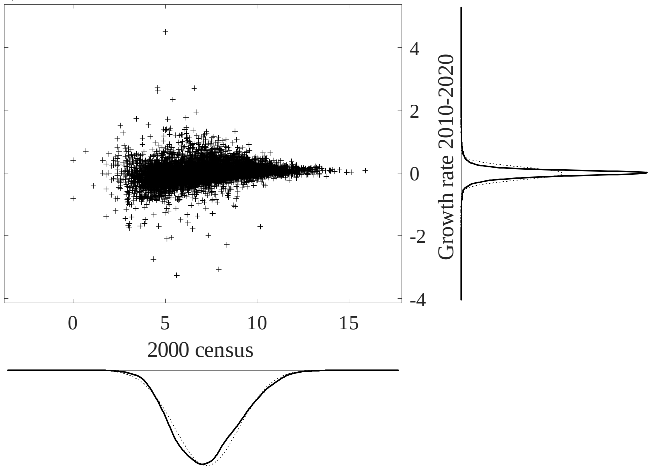
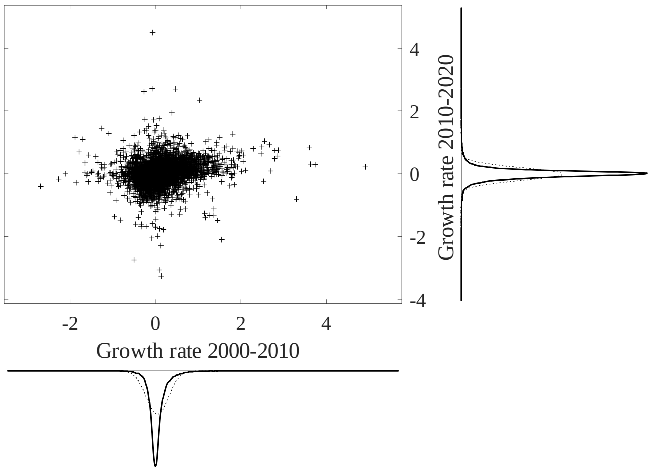


Figure 1c: Growth rates between 2000 and 2010, and between 2010 and 2020



Notes: Scatter plot of (log) city sizes for the contiguous US states in 2000 and their (continuously compounded) growth rates between 2000 and 2010, and 2010 and 2020, as well as kernel density estimates of their marginal distributions (continuous lines), together with the best normal approximation to them (dotted lines), which share their sample means and standard deviations. Sample: 23,830 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; see Section 5 for details.