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

Abstract

We provide a pedagogical proof that the information matrix test for a multivariate normal random vector coincides with the sum of the two moment tests that look at the means of all the different third- and fourth-order multivariate Hermite polynomials evaluated at the sample mean and covariance matrix. We also show that its finite sample distribution does not depend on either the true values of the mean vector and covariance matrix or their sample counterparts, so it can be obtained to any degree of accuracy by simulating spherical Gaussian vectors and creating orthogonalized residuals using the first two sample moments.

Keywords: Exact test, Hessian matrix, Multivariate normality, Outer product of the score.

JEL: C30, C46, C52

*The first and third authors acknowledge financial support from the Spanish Ministry of Economy, Industry & Competitiveness through grant ECO 2017-89689 and the Santander CEMFI Research Chair.

1 Introduction

The information matrix test introduced by White (1982) constitutes a rather general procedure for examining the specification of models estimated by maximum likelihood (ML). As is well known, it directly assesses the information matrix equality, which states that the sum of the Hessian matrix and the outer product of the score vector should be 0 in expected value when the estimated model is correctly specified. As an illustration, White (1982) looked at the information matrix test for a univariate normal random variable, which simply checks that the third- and fourth-order Hermite polynomials of the standardised variable have 0 means in the population. Therefore, it is equivalent to the version of the popular Jarque and Bera (1980) test proposed by Kiefer and Salmon (1981) among many others. In this note, we show that the information matrix 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, respectively, thereby generalising the univariate result.

There is an extensive literature in econometrics on the theoretical properties and interpretation of the information matrix test, as well as on its applications and finite sample behaviour. There is also a huge literature in statistics on multivariate normality tests. To the best of our knowledge, though, the intersection is void. Given the univariate precedent, it is not surprising that the information matrix test statistic is 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 information matrix test in Chesher (1984) provides a completely different justification, which might be more relevant in some empirical applications.

Importantly, we explicitly address the widespread and often justified concern that the information matrix is unreliable in finite samples by explaining how to simulate its exact, parameter-free, finite sample distribution to any desired degree of accuracy for any dimension of the random vector and sample size. In this respect, we exploit the numerical invariance of the test statistic to affine transformations of the observed variables to simulate draws extremely quickly.

The rest of the note is organised as follows. We include our theoretical results in section 2 and discuss computational issues in section 3. Next, we present the results of some Monte Carlo exercises looking at the power of the test in finite samples, and finish by mentioning some avenues for further research. Proofs and auxiliary results are relegated to appendices.

2 The information matrix test

Our null hypothesis is that

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

with $\boldsymbol{\nu}, \boldsymbol{\Gamma}$ unknown.[1] Let $\boldsymbol{\Delta} = \boldsymbol{\Gamma}^{-1}$ and $\boldsymbol{\varepsilon}(\boldsymbol{\nu}) = (\mathbf{x} - \boldsymbol{\nu})$. Barndorff-Nielsen and Petersen (1979) define the (centred) multivariate Hermite polynomials of $\mathbf{x}$ of order $k = k_1 + \ldots + k_N \geq 0$ as

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

As is well known, when model (1) is correctly specified, the mean of any Hermite polynomial of positive degree is 0. We can then state our main result:

Proposition 1 *The information matrix test that 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 that check whether the expected values of all the distinct third- and fourth-order multivariate Hermite polynomials of* $\mathbf{x}_t$ *are 0.*

Following Chesher (1984), we can interpret the moment test of the fourth-order multivariate Hermite polynomials as a test of neglected heterogeneity in the covariance matrix of the observations. Similarly, the test that looks at the third-order ones effectively assesses dependence in the neglected heterogeneity of the mean and covariance parameters. In contrast, neglected heterogeneity in the vector of mean parameters is untestable because the means of the second-order multivariate Hermite polynomials are always 0 when the covariance matrix $\boldsymbol{\Gamma}$ is freely estimated.

Multivariate Hermite polynomials of different orders are uncorrelated (see Holmquist (1996)), which justifies the additive decomposition of the test statistic in Proposition 1. Following the lead of Phillips and Park (1988), Holly and Gardiol (1995) explain how to obtain matrix expressions for the covariance matrices of the entire vector of polynomials of any given common order. But the symmetry of the higher-order partial derivatives in (2) implies that some of the N^k multivariate Hermite polynomials of order k will be replicated several times. Specifically, there are only $\binom{N+k-1}{k}$ different polynomials, so we can avoid generalised inverse matrices by eliminating the redundant ones 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. Thus, we end up with $N(N+1)(N+2)/6$ and $N(N+1)(N+2)(N+3)/24$ distinct third- and fourth-order moment conditions, respectively, which coincide with the degrees of freedom of the asymptotic chi-square distributions of the corresponding multivariate

[1] If $\boldsymbol{\nu}$ and $\boldsymbol{\Gamma}$ were known, there would be no parameters to estimate under the null, and therefore no gradient or information matrix. Still, the test statistic in Proposition 1 would continue to be valid as a multivariate normality test. Similarly, we use the *i.i.d.* asumption mainly for computing the asymptotic variance of the influence functions, which in principle could be robustified for the presence of serial correlation.

skewness and kurtosis tests. The next result contains detailed expressions for the covariances between two arbitrary third- and fourth-order polynomials:

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

$$cov(H_{ijk}, H_{i'j'k'}) = \delta_{ii'}\delta_{jj'}\delta_{kk'} + \delta_{ii'}\delta_{jk'}\delta_{kj'} + \delta_{ij'}\delta_{ji'}\delta_{kk'} + \delta_{ij'}\delta_{jk'}\delta_{ki'} + \delta_{ik'}\delta_{ji'}\delta_{kj'} + \delta_{ik'}\delta_{jj'}\delta_{ki'},$$

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

When $\mathbf{\Gamma} = \mathbf{I}_N$, the components of $\mathbf{x}_t$ are stochastically independent and the multivariate Hermite polynomial $H_{1^{k_{11}}\ldots N^{k_{NN}}}[\boldsymbol{\varepsilon}(\boldsymbol{\nu}), \mathbf{\Delta}]$ simplifies to the product of the univariate polynomials $H_{1^{k_{11}}}[\varepsilon_1(\nu_1)]\ldots H_{N^{k_{NN}}}[\varepsilon_N(\nu_N)]$. Moreover, Lemma 1 implies that different multivariate Hermite polynomials of the same order become orthogonal to each other, so the information matrix 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 $\binom{N+2}{3} + \binom{N+3}{4}$ asymptotically independent moment tests, which would be easy to compute.

3 Computational considerations

Consider the following full-rank affine transformation $\mathbf{y}_t = \mathbf{a} + \mathbf{B}\mathbf{x}_t$ with $|\mathbf{B}| \neq 0$. As is well known, $\mathbf{y}_t \sim i.i.d.\ N(\mathbf{a}+\mathbf{B}\boldsymbol{\nu}, \mathbf{B}\mathbf{\Gamma}\mathbf{B}')$ when (1) holds. Our next result shows that the information matrix test statistic is numerically invariant to the values of $\mathbf{a}$ and $\mathbf{B}$:

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

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 T on $\mathbf{x}_t$, 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 $\mathbf{I}_N$, respectively. Thus, the information matrix 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 $\mathbf{\Gamma}$ by applying the same trick.

Proposition 2 also implies that the sample mean vector and covariance matrix of the observations, which set to 0 the average of the first and second multivariate Hermite polynomials, 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 $\mathbf{x}_t$ and sample size. In fact, it suffices to simulate R times a random sample of size T of a spherical Gaussian random vector of dimension N to obtain R independent draws of the information matrix test statistic for multivariate normality. Although this can be regarded as a parametric bootstrap procedure that provides the exact p-value of the test statistic obtained in a real sample as the number of bootstrap replications R grows without bound, the fact that the only characteristics of the original sample that matter are the values of N and T implies that a researcher could obtain tables with exact critical values before observing the data.

Given that the sample mean and covariance matrix of a multivariate random vector take hardly any time to compute, and that the information matrix 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.

4 Monte Carlo evidence

The discussion in the previous section indicates that assessing the finite sample size of our proposed test only makes sense if R were small. For that reason, in this section we focus on the small sample power of the information matrix test by means of an extensive Monte Carlo simulation exercise in which we generate 20,000 samples from three multivariate non-Gaussian distributions whose mean vector and covariance matrix are $\mathbf{0}$ and $\mathbf{I}_N$, respectively: the asymmetric Student t distribution and 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 (1987) test in the bivariate case.

We make use of Proposition 2 not only in fixing the population mean vector and covariance matrix, which are nevertheless freely estimated in the sample, but also in exploiting that for these three distributions skewness is a common feature (see Engle and Kozicki (1993)), so that one can always find orthogonal rotations of the original random vectors in which only one variable is asymmetric. 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. The main difference between the skew normal distribution and the other two, though, is that in the former the other $N-1$ variables are Gaussian and independent, so that all the remaining third and fourth multivariate cumulants are 0, while in the latter, those variables are symmetric but neither normal nor independent. 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.[2]

We accurately approximate the finite sample critical values with $R = 10^6$ replications and report the rejection rates at the 5% significance level for three dimensions ($N = 2, 4, 8$) and three sample lengths ($T = 64, 256, 1024$) in Table 1. As expected, power increases with the sample size T. Similarly, power increases with N for the two LSMNs but it decreases for the skew normal. The reason is simple. Given the canonical representation of the skew normal mentioned above, the only thing that increasing N does is to add more independent Gaussian components, which in turn add more 0 (co-)skewness and (co-)kurtosis terms. As a result, the non-centrality parameter does not change while the number of degrees of freedom increases.

5 Directions for future research

Our Monte Carlo exercises confirm the non-trivial power of the information matrix test against empirically plausible alternatives, even though it is not consistent because in arbitrary large samples it would fail to reject with probability one departures from normality such that all third- and fourth-order cumulants are 0. Unlike in the univariate case, though, it is not obvious how to construct distributions with this characteristic because it is difficult to ensure the global positivity of multivariate Hermite expansions of the Gaussian density.

The information matrix test could be extended to multivariate, conditionally heteroskedastic, dynamic regression models with Gaussian innovations, but the number of moments involved would increase very quickly. Nevertheless, our proposed test continues to be asymptotically valid as a test of multivariate normality for the standardised residuals of such a model when its parameters have been estimated under the null because the Gaussian scores are conditionally linear transformations of the first and second multivariate Hermite polynomials of the innovations. Unfortunately, exact finite sample distributions only seem feasible for the multivariate linear regression model with fixed regressors analysed by Dufour, Khalaf and Beaulieu (2003). In more general models, though, a parametric bootstrap would usually offer a higher-order approximation to the finite sample distribution.

The information matrix test could also be extended to examine the specification of more general multivariate distributions. We are currently exploring some of these interesting research avenues.

[2] The kurtosis of a skew normal is a function of its skewness parameter only (see Suplemental Appendix C).

References

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.

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

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

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.

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

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.

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.

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

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

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.

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

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

Appendix

Proof of Proposition 1

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

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

where $\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma}) = \mathbf{\Delta}\boldsymbol{\varepsilon}(\boldsymbol{\nu}) = \mathbf{\Gamma}^{-1}(\mathbf{x}-\boldsymbol{\nu})$ and $\boldsymbol{\gamma} = vech(\mathbf{\Gamma})$. 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

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

which coincide with the product of the (transposed) duplication matrix $\mathbf{D}_N$ and the second-order Hermite polynomials.

Therefore, the Hessian matrix is given by

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

and

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

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

$$\begin{aligned}
\boldsymbol{\nu}\boldsymbol{\nu} &: \ \mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - \mathbf{\Delta} \\
\boldsymbol{\gamma}\boldsymbol{\nu} &: \ \frac{1}{2}\mathbf{D}_N'\{vec[\mathbf{\Delta}\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - \mathbf{\Delta}]\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - 2[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) \otimes \mathbf{\Delta}]\}
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\gamma}\boldsymbol{\gamma} \ : \ & \frac{1}{4}\mathbf{D}_N' vec[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma}) - \mathbf{\Delta}]vec'[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{\Delta} - \mathbf{\Delta}]\mathbf{D}_N \\
& -\frac{1}{2}\mathbf{D}_N'\{2[\mathbf{z}(\boldsymbol{\nu},\boldsymbol{\gamma})\mathbf{z}'(\boldsymbol{\nu},\boldsymbol{\gamma})] - (\mathbf{\Delta} \otimes \mathbf{\Delta})\}\mathbf{D}_N.
\end{aligned}$$

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

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

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

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

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

$$vec[vec(\mathbf{z}\mathbf{z}' - \boldsymbol{\Delta})\mathbf{z}' - 2(\mathbf{z} \otimes \boldsymbol{\Delta})] = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}) - (\mathbf{z} \otimes \boldsymbol{\delta}) - 2(\mathbf{K}_{NN} \otimes \mathbf{I}_N)(\mathbf{z} \otimes \boldsymbol{\delta}), \tag{1}$$

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

$$\begin{aligned} vec[vec(\mathbf{z}\mathbf{z}')\mathbf{z}'] &= [\mathbf{z} \otimes vec(\mathbf{z}\mathbf{z}')] = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec[vec(\boldsymbol{\Delta})\mathbf{z}'] &= (\mathbf{z} \otimes \boldsymbol{\delta}), \\ vec(\mathbf{z} \otimes \boldsymbol{\Delta}) &= (1 \otimes \mathbf{K}_{NN} \otimes \mathbf{I}_N)(\mathbf{z} \otimes \boldsymbol{\delta}) = (\mathbf{K}_{NN} \otimes \mathbf{I}_N)(\mathbf{z} \otimes \boldsymbol{\delta}), \end{aligned}$$

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

Finally, the $\boldsymbol{\gamma\gamma}$ block will be

$$\begin{gathered} vec\{vec(\mathbf{z}\mathbf{z}' - \boldsymbol{\Delta})vec'(\mathbf{z}\mathbf{z}' - \boldsymbol{\Delta}) - [4(\boldsymbol{\Delta} \otimes \mathbf{z}\mathbf{z}') - 2(\boldsymbol{\Delta} \otimes \boldsymbol{\Delta})]\} \\ = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}) - (\mathbf{z} \otimes \mathbf{z} \otimes \boldsymbol{\delta}) - 5(\boldsymbol{\delta} \otimes \mathbf{z} \otimes \mathbf{z}) + (\boldsymbol{\delta} \otimes \boldsymbol{\delta}) + 2(\mathbf{I}_N \otimes \mathbf{K}_{NN} \otimes \mathbf{I}_N)(\boldsymbol{\delta} \otimes \boldsymbol{\delta}) \end{gathered} \tag{2}$$

because

$$\begin{aligned} vec[vec(\mathbf{z}\mathbf{z}')vec'(\mathbf{z}\mathbf{z}')] &= [vec(\mathbf{z}\mathbf{z}') \otimes vec(\mathbf{z}\mathbf{z}')] = (\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec[\boldsymbol{\delta} vec'(\mathbf{z}\mathbf{z}')] &= [vec(\mathbf{z}\mathbf{z}') \otimes \boldsymbol{\delta}] = (\mathbf{z} \otimes \mathbf{z} \otimes \boldsymbol{\delta}), \\ vec[vec(\mathbf{z}\mathbf{z}')\boldsymbol{\delta}'] &= [\boldsymbol{\delta} \otimes vec(\mathbf{z}\mathbf{z}')] = (\boldsymbol{\delta} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec(\boldsymbol{\delta}\boldsymbol{\delta}') &= (\boldsymbol{\delta} \otimes \boldsymbol{\delta}), \\ vec(\boldsymbol{\Delta} \otimes \mathbf{z}\mathbf{z}') &= (\mathbf{I}_N \otimes \mathbf{K}_{1N} \otimes \mathbf{I}_N)[\boldsymbol{\delta} \otimes vec(\mathbf{z}\mathbf{z}')] = (\boldsymbol{\delta} \otimes \mathbf{z} \otimes \mathbf{z}), \\ vec(\boldsymbol{\Delta} \otimes \boldsymbol{\Delta}) &= (\mathbf{I}_N \otimes \mathbf{K}_{NN} \otimes \mathbf{I}_N)(\boldsymbol{\delta} \otimes \boldsymbol{\delta}) \end{aligned}$$

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

$$\begin{aligned} &\mathbf{S}_{N\boldsymbol{\iota}_1}\mathbf{z} \\ &\mathbf{S}_{N\boldsymbol{\iota}_2}[(\mathbf{z} \otimes \mathbf{z}) - \boldsymbol{\delta}], \\ &\mathbf{S}_{N\boldsymbol{\iota}_3}[(\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}) - 3(\boldsymbol{\delta} \otimes \mathbf{z})], && (3) \\ &\mathbf{S}_{N\boldsymbol{\iota}_4}[(\mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z} \otimes \mathbf{z}) - 6(\mathbf{z} \otimes \mathbf{z} \otimes \boldsymbol{\delta}) + 3(\boldsymbol{\delta} \otimes \boldsymbol{\delta})], && (4) \end{aligned}$$

where $\mathbf{S}_{N\boldsymbol{\iota}_k}$ $(k = 1, \ldots, 4)$ are the symmetrisation operators discussed by Homlquist (1996), whose detailed expressions we provide in Supplemental Appendix A.

It is easy to see that the $\boldsymbol{\nu\nu}$ term coincides with the second-order Hermite polynomials because $\mathbf{S}_{N\boldsymbol{\iota}_2}$ applied to $(\mathbf{z}\otimes\mathbf{z})$ has no effect and $\mathbf{K}_{NN}\boldsymbol{\delta} = \boldsymbol{\delta}$ 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, $\mathbf{S}_{N\boldsymbol{\iota}_3}$ applied to $(\mathbf{z}\otimes\mathbf{z}\otimes\mathbf{z})$ has no effect either. In contrast, if we apply $6\mathbf{S}_{N\boldsymbol{\iota}_3}$ to $(\mathbf{z}\otimes\boldsymbol{\delta})$ we obtain

$$
\begin{gathered}
[\mathbf{I}_{N^3} + (\mathbf{I}_N\otimes\mathbf{K}_{NN}) + (\mathbf{K}_{NN}\otimes\mathbf{I}_N) + (\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N) \\
+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN}) + (\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N)](\mathbf{z}\otimes\boldsymbol{\delta}) \\
= (\mathbf{z}\otimes\boldsymbol{\delta}) + (\mathbf{z}\otimes\boldsymbol{\delta}) + (\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta}) + (\boldsymbol{\delta}\otimes\mathbf{z}) \\
+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta}) + (\boldsymbol{\delta}\otimes\mathbf{z}) \\
= 2[(\mathbf{z}\otimes\boldsymbol{\delta}) + (\boldsymbol{\delta}\otimes\mathbf{z}) + (\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})],
\end{gathered}
$$

so that

$$
\begin{aligned}
(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{z}\otimes\boldsymbol{\delta}) &= (\mathbf{z}\otimes\boldsymbol{\delta}), \\
(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta}) &= \mathbf{K}_{N^2N}(\mathbf{z}\otimes\boldsymbol{\delta}) = (\boldsymbol{\delta}\otimes\mathbf{z})
\end{aligned}
$$

by virtue of Theorems 3.7 (iii) and 3.1 in Magnus (1986), and

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

Similarly,

$$
\begin{gathered}
6\mathbf{S}_{N\boldsymbol{\iota}_3}(\boldsymbol{\delta}\otimes\mathbf{z}) = [\mathbf{I}_{N^3} + (\mathbf{I}_N\otimes\mathbf{K}_{NN}) + (\mathbf{K}_{NN}\otimes\mathbf{I}_N) + (\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N) \\
+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN}) + (\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N)](\boldsymbol{\delta}\otimes\mathbf{z}) \\
= (\boldsymbol{\delta}\otimes\mathbf{z}) + (\mathbf{I}_N\otimes\mathbf{K}_{NN})(\boldsymbol{\delta}\otimes\mathbf{z}) + (\boldsymbol{\delta}\otimes\mathbf{z}) + (\mathbf{I}_N\otimes\mathbf{K}_{NN})(\boldsymbol{\delta}\otimes\mathbf{z}) \\
+(\mathbf{z}\otimes\boldsymbol{\delta}) + (\mathbf{z}\otimes\boldsymbol{\delta}) \\
= 2[(\mathbf{z}\otimes\boldsymbol{\delta}) + (\boldsymbol{\delta}\otimes\mathbf{z}) + (\mathbf{I}_N\otimes\mathbf{K}_{NN})(\boldsymbol{\delta}\otimes\mathbf{z})],
\end{gathered}
$$

because

$$
\begin{aligned}
(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\boldsymbol{\delta}\otimes\mathbf{z}) &= (\boldsymbol{\delta}\otimes\mathbf{z}), \\
(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\boldsymbol{\delta}\otimes\mathbf{z}) &= \mathbf{K}_{NN^2}(\boldsymbol{\delta}\otimes\mathbf{z}) = (\mathbf{z}\otimes\boldsymbol{\delta})
\end{aligned}
$$

by virtue of expression (3.3) in Magnus (1986), which implies that $\mathbf{K}_{NN^2} = \mathbf{K}_{N^2N}^{-1}$, and his Theorem 3.1.

Finally,

$$6\mathbf{S}_{N\boldsymbol{\iota}_3}(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta}) = [\mathbf{I}_{N^3}+(\mathbf{I}_N\otimes\mathbf{K}_{NN})+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)+(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N)$$

$$+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN})+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{I}_N\otimes\mathbf{K}_{NN})(\mathbf{K}_{NN}\otimes\mathbf{I}_N)](\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})$$

$$=(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{z}\otimes\boldsymbol{\delta})+(\mathbf{z}\otimes\boldsymbol{\delta})$$

$$+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})$$

$$=2[(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})].$$

because

$$(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{K}_{NN}\otimes\mathbf{I}_N)=\mathbf{I}_{N^3}.$$

Hence,

$$\begin{aligned} &\mathbf{S}_{N\boldsymbol{\iota}_3}[(\mathbf{z}\otimes\boldsymbol{\delta})+2(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})] \\ = \quad &[(\mathbf{z}\otimes\boldsymbol{\delta})+(\boldsymbol{\delta}\otimes\mathbf{z})+(\mathbf{K}_{NN}\otimes\mathbf{I}_N)(\mathbf{z}\otimes\boldsymbol{\delta})] = 3\mathbf{S}_{N\boldsymbol{\iota}_3}(\mathbf{z}\otimes\boldsymbol{\delta}),\end{aligned}$$

so that (1) does indeed coincide with (3).

A very tedious but entirely analogous procedure confirms that $\mathbf{S}_{N\boldsymbol{\iota}_4}$ applied to (2) coincides with (4). □

Proof of Lemma 1

The proof is a careful but straightforward implemented using a computer algebra system of the procedure described in Holly and Gardiol (1995), who rely on the formulas for the higher order moments of the multivariate normal in Balestra and Holly (1990), which in turn generalised Magnus and Neudecker (1979) and Phillips and Park (1988). □

Proof of Proposition 2

Given that the mapping from $\mathbf{x}_t$ to $\mathbf{y}_t$ 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 $\mathbf{y}_t$ will be $\mathbf{B}^{\otimes k} = \underbrace{\mathbf{B}\otimes\mathbf{B}\otimes\ldots\otimes\mathbf{B}}_{k\text{ times}}$ times the vector of multivariate Hermite polynomials of order k for $\mathbf{x}_t$. The numerical invariance of moment tests to linear transformations of the influence functions with constant coefficients yields the desired result. □

Table 1: Monte Carlo rejection rates at the 5% significance level

Panel A: Joint test of (co-)skewness components

		Asymmetric t			Mixture of normals			Skew normal		
$N \setminus T$	*df*	64	256	1,024	64	256	1,024	64	256	1,024
2	*4*	44.42	96.17	100.00	64.52	98.45	100.00	37.44	97.90	100.00
4	*20*	45.86	96.72	100.00	83.30	99.82	100.00	17.75	80.80	100.00
8	*120*	55.67	98.56	100.00	98.80	100.00	100.00	8.935	38.06	99.78

Panel B: Joint test of (co-)kurtosis components

		Asymmetric t			Mixture of normals			Skew normal		
$N \setminus T$	*df*	64	256	1,024	64	256	1,024	64	256	1,024
2	*5*	31.77	69.41	99.12	58.21	97.61	100.00	14.78	30.57	69.53
4	*35*	39.06	80.91	99.87	84.09	99.97	100.00	9.62	19.47	48.37
8	*330*	56.48	96.77	100.00	99.41	100.00	100.00	7.21	11.69	25.34

Panel C: Joint test of (co-)skewness and (co-)kurtosis components

		Asymmetric t			Mixture of normals			Skew normal		
$N \setminus T$	*df*	64	256	1,024	64	256	1,024	64	256	1,024
2	*9*	38.93	91.52	100.00	66.12	99.59	100.00	24.01	88.25	100.00
4	*55*	43.15	92.19	100.00	87.27	99.99	100.00	12.07	46.47	99.99
8	*450*	58.06	98.40	100.00	99.50	100.00	100.00	7.59	17.64	81.86

Notes: *df* denotes degrees of freedom. 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 $\mathbf{I}_N$, 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 C for details.

Supplemental Appendix for

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

A The symmetrisation operators

The first four symmetrisation operators discussed by Homlquist (1996) are

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

which applied to the arbitrary vectors of dimension N $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ yield

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

B Special cases

B.1 The univariate case

The contribution of x to the log-likehood 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

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

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

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

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

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

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 0. However, the expected value of the first term will also be 0 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 The bivariate case

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

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

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

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

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 $\boldsymbol{\nu}\boldsymbol{\nu}$ term of the sum of the outer product of the score and the Hessian matrix are identical to these polynomials.

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

$$\left[\begin{matrix} (\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11})(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) \\ (\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) \end{matrix}\right.$$
$$\left.\begin{matrix} (\delta_{11}^2\varepsilon_1^2 + 2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2 + \delta_{12}^2\varepsilon_2^2 - \delta_{11})(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) \\ (\delta_{11}\delta_{12}\varepsilon_1^2 + (\delta_{12}^2 + \delta_{11}\delta_{22})\varepsilon_1\varepsilon_2 + \delta_{22}\delta_{12}\varepsilon_2^2 - \delta_{12})(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) \\ (\delta_{12}^2\varepsilon_1^2 + 2\delta_{12}\delta_{22}\varepsilon_1\varepsilon_2 + \delta_{22}^2\varepsilon_2^2 - \delta_{22})(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) \end{matrix}\right]$$
$$-2\begin{bmatrix} \delta_{11}(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) & \delta_{12}(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) \\ \delta_{12}(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) & \delta_{22}(\delta_{11}\varepsilon_1 + \delta_{12}\varepsilon_2) \\ \delta_{11}(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) & \delta_{12}(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) \\ \delta_{12}(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) & \delta_{22}(\delta_{12}\varepsilon_1 + \delta_{22}\varepsilon_2) \end{bmatrix},$$

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 $N(N+1)(N+2)/6 = 4$ in the bivariate case. Those four terms are

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

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

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

and

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

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}^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}^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_{11}^2\varepsilon_1^2+2\delta_{11}\delta_{12}\varepsilon_1\varepsilon_2+\delta_{12}^2\varepsilon_2^2-\delta_{11}) & 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_{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})(\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_{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}) & (\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}.
$$

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{gathered}
(\mathbf{1},\mathbf{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 \\
-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},\mathbf{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 \\
-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 \\
-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},\mathbf{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 \\
-\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
\end{gathered}$$

$$\begin{gathered}
(\mathbf{1},\mathbf{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} \\
+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 \\
-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},\mathbf{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 \\
-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} \\
-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},\mathbf{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} \\
+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}
\end{gathered}$$

$$\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} \\
(\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 $N(N+1)(N+2)(N+3)/24 = 5$ in the bivariate case. Those five terms are

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

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

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

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

and

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

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

C Alternative distributions

For the multivariate skew normal distribution, we use its canonical representation, choosing .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 3.596 for its skewness and kurtosis coefficients (see Figure 2.2 in Azzalini and Capetiano (2014) for the feasible skewness-kurtosis combinations). In contrast, the remaining $N-1$ components are drawn from independent univariate standard normals.

In the case of the multivariate asymmetric Student t, we choose $\eta = .042$ and $\mathbf{b} = (-.91, \mathbf{0}')'$, 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 $\mathbf{I}_N$, respectively). Finally, for the discrete mixture of two normal vectors, we fix their means to $(1-\lambda)\boldsymbol{\delta}$ and $-\lambda\boldsymbol{\delta}$, where $\lambda = 1/4$ is the probability of the first Gaussian vector and $\boldsymbol{\delta} = (-.57, \mathbf{0}')'$, and their covariance matrices to

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

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.